DOCUMENT RESUME

ED 144 611 IR 005 259

AUTHOR Persson, Olle; Hoglund, Lars
TITLE Evaluation of a Computer-Based Current Awareness Service for Swedish Social Scientists. Research Report No. 29.
INSTITUTION Royal Inst. of Tech., Stockholm (Sweden). Library.; Umea Univ. (Sweden).
REPORT NO TRITA-LIB-6003
PUB DATE Mar 75
NOTE 84p.; Report developed by the Department of Sociology, Umea University

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS *Citation Indexes; Data Bases; Evaluation; *Information Dissemination; *Information Retrieval; *Information Science; *Social Sciences
IDENTIFIERS *Selective Dissemination of Information; *Social Science Citation Index

ABSTRACT

This report presents results from an evaluation of a current awareness service for Swedish social scientists. The service was based on the computerized version of Social Science Citation Index (SSCI). For the study an evaluation model was developed where the interaction between user and system is evaluated according to specific criteria for each phase of the information-retrieval process. Generally, system performance during the thirteen week test-period was good compared to other data-bases. However, the evaluation suggests various means of improving the service. One of the most crucial factors affecting system performance and user satisfaction seems to be user knowledge about retrieval functions. Clarity and completeness of queries, degree of profiling activity, and coverage of the data base were other factors. While output was somewhat low, the mean precision rate was satisfactory. Users' evaluation of the SSCI usefulness showed about 65% of the 500 users considered the SSCI to be useful to them. The general value of SSCI is indicated by the fact that, during the rather short test-period, a majority of users increased their knowledge about current research and became acquainted with journals which they did not know previously. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED144611

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

RESEARCH REPORTS FROM THE
DEPARTMENT OF SOCIOLOGY
UNIVERSITY OF UMEÅ

NO 29 MARCH 1975

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stephan Schwarz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

OLLE PERSSON AND LARS HÖGLUND

EVALUATION OF A COMPUTER-BASED CURRENT AWARENESS SERVICE FOR SWEDISH SOCIAL SCIENTISTS [1)]

(Social Science Citation Index)

1) This report was first published in Stockholm Papers in Library and Information Science, TRITA-LIB-6003

IR005 259

ABSTRACT

This report presents results from an evaluation study of a current awareness service for Swedish social scientists. The service was based on the computerized version of Social Science Citation Index (SSCI). For the study an evaluation model was developed where the interaction between user and system is evaluated according to specific criteria for each phase of the IR-process.

Generally, system performance during the thirteen week test-period was rather good compared to other data-bases. However, the evaluation study suggests various means of improving the service. One of the most crucial factors affecting system performance and user satisfaction seems to be user knowledge about IR-functions. Other crucial factors were clarity and completeness of queries, degree of profiling activity and coverage of the data-base.

While output was somewhat low, the mean precision rate was quite satisfactory. Users' overall evaluation of the SSCI-usefulness showed that about 65 per cent of the 500 users considered the SSCI to be useful to them. The general value of SSCI is indicated by the fact that during the rather short test-period a majority of the users increased their knowledge about current research and became aquainted with journals which they did not know of previously. Finally, the SSCI seems to have a broad group of potential users among professionals at universities, research institutes, government agencies and industry.

CONTENTS

		Page
I	SSCI AND PATTERNS OF INFORMATION USE ON SOCIAL SCIENCE	1
	1.1 Introduction	1
	1.2 The idea of computerized information retrieval	2
	1.3 User attitudes towards current awareness	3
	1.4 The potential of SSCI for information retrieval	5
II	EVALUATING INFORMATION RETRIEVAL SYSTEMS	10
	2.1 Some general evaluation problems	10
	2.2 Phases and criteria of the computerized IR-process	13
	2.3 Summary of the evaluation model	17
III	DESIGN OF THE EVALUATION STUDY	19
	3.1 Definition of the studied population of SSCI-users	19
	3.2 Pre-test survey - before the test period	19
	3.3 Evaluation of reference lists - during the test period	20
	3.4 After the test period	20
	3.5 Comments on the data gathered	21
	3.6 Comments on the methods used in relation to the evaluation strategy	23
IV	INTRODUCTION OF THE SSCI	25
	4.1 The user group	25
	4.2 Information about the aims and functions of the SSCI-service	29
V	QUERY FORMULATION	32
	5.1 Degree of clarity in queries	33
	5.2 Extra information in queries	34
	5.3 The correspondence of queries and data-base	34
VI	PROFILING	37
	6.1 Profiling activity	37
	6.2 Query-Profile correspondence	39

VII SYSTEM OUTPUT AND USER EVALUATION 41
7.1 System output 41
7.2 User evaluation 44

VIII FACTORS AFFECTING SYSTEM PERFORMANCE 49
8.1 Selecting a measure for system performance 49
8.2 Introduction phase 51
8.3 Query formulation 52
8.4 Profiling 53
8.5 User characteristics 55
8.6 Inter-relations between some criteria of the IR-process 57

IX SUMMARY AND CONCLUSIONS 62
9.1 Improving the different phases of the IR-process 62
9.2 Further marketing of the SSCI 66
9.3 The general value of SSCI for social science research and practice 67

X REFERENCES

XI APPENDICES 73
I - Letter of introduction of the SSCI 73
II - Example of query 75
III:1-4 - Examples of search profiles 77
IV - Example of reference list 81
V - Example of user's interest evaluation 82
VI - Source journals of the SSCI 83

PREFACE

Computer-based information retrieval systems (IR-systems) are now available in a number of fields in science and technology. During the last few years some data-bases for social science literature have also been designed and made available (ERIC for educational research, PA for phychological research and now SSCI for journals covering several areas of social science).

This is the first report on an evaluation study of a computerized information service for Swedish social scientists based on the Social Science Citation Index (SSCI). The information service based on SSCI has been designed as a current awareness service with selective dissemination of information according to individual search profiles (SDI). The service is administered by the Library at the Royal Institute of Technology in Stockholm, where the Information and Documentation Center (IDC) has, since 1967, been engaged in research on computer-based information retrieval with financial support from the Swedish Council for Scientific Information and Documentation (SINFDOK).

In 1973 the IDC accepted an offer from the Institute for Scientific Information (ISI) to test the SSCI-tapes, which cover about 1 000 journals in the social sciences. The intention was to introduce this information system free of charge in order to get some impressions of the possibilities of computerized information retrieval services for the social sciences. This implied that an evaluation study would be conducted. For the evaluation of the SSCI-service the IDC received a grant from the Swedish Council for Social Science Research.

The evaluation study has been conducted at the Department of Sociology, University of Umeå by Lars Höglund and Olle Persson in close cooperation with the IDC and with Roland Hjerppe and Zofia Gluchowich in particular, whose support has been indispensable. We want to thank Ola Román for serving as our major advisor at the department of Sociology. We also want to thank Georg Karlsson for his constructive suggestions and encouragement. We are grateful to Patricia Shrimpton for her patient correction of linguistic errors.

I SSCI AND PATTERNS OF INFORMATION USE IN SOCIAL SCIENCE

1.1 Introduction

Ever since the first scientific journals were published in 1665, the number of such journals has increased at an exponential rate and today the scientific journal is a major channel for the dissemination of research results among scientists. However, nowadays the enormous number of journals makes it very difficult to keep up with the rapid development in many fields. Aids such as abstract-journals and bibliographies, are often too slow to keep up with the research front. Consequently, we have seen a number of new literature services introduced over the last 15 years. The most significant of these new services are the computer-based information retrieval systems (IR-systems).

The major advantage of computerized information retrieval is its capacity to select items of information from a large stock according to specific information needs. The aim of an evaluation of an IR-system then is principally to study how well this selection process is carried out. The IR-process involves several steps or phases which have to be passed through in order to reach the final retrieval of references. To be of practical value an evaluation study should try to detect the causes of system performance in each phase of the IR-process. This is therefore the approach chosen for this evaluation study.

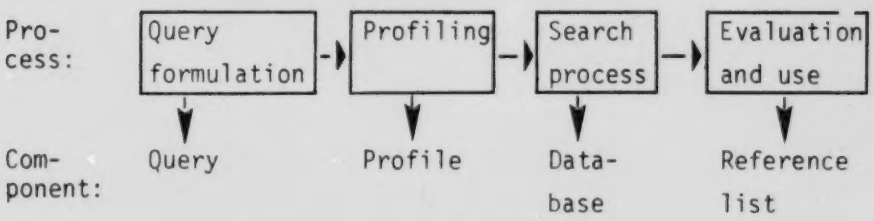
The plan of this report follows logically from this kind of evaluation approach. In this chapter we describe the general idea of computerized information retrieval and discuss the potential of SSCI as an information source for social scientists. In Chapter II we discuss some general problems in evaluating information retrieval systems and present a strategy for such evaluations. The design of the evaluation study of SSCI is outlined in Chapter III. In Chapters IV-VII the different phases

of the IR-process are evaluated: The introductory phase, query formulation, profiling, system output and user evaluation. Chapter VIII deals with the analyses of the causes of system performance. In Chapter IX the evaluation study is summarized and practical actions for improving the service are suggested. Finally, we discuss the further marketing of SSCI in Sweden and the general value of SSCI for social science research.

1.2 The idea of computerized information retrieval

The data-based SSCI covers about 1 000 scientific journals in different fields of social science and is continuously issued each week. This means that each issue of the data-base contains only new articles and is designed primarily as a current awareness service.

Fig. 1.1 The processes and components of computerized information retrieval



As Fig. 1.1 indicates, the computerized information retrieval can be looked upon as a process or as a system of related components. Vickery (1973) defines an information system as (1) a system of interlinked entities and (2) a system of interlinked processes. In reality the entities participate in the process and their interaction is indicated in Fig. 1.1.

The starting point in the IR-process is the query formulation, i.e. when users make an explicit verbalization of their information needs (query). Profiling is the process by which the query is

transformed into a search profile. The search process mechanically retrieves references from the data-base according to search terms and the search logic of the profile. In this case the searchable elements of the data-base are title of journal article, author and journal name. Each search results in a printed list of references, which is sent to users for evaluation and use. The IDC can use the evaluations for necessary profile modifications.

The data-base SSCI is primarily designed as a current awareness service. The need for such a service is emphasized by the ever-increasing volume of journals in all languages, the growing interdisciplinary character of most fields of study and the explosive growth of technical reports, preprints etc. which means that current awareness in even a limited subject field is a problem of great magnitude. In this respect the SDI-service based on the SSCI has some important advantages that can be mentioned.

The data-base has an extensive coverage, which permits inter-disciplinary searches through the social sciences. The data-base is also complete by up-to-date since it is issued on a weekly basis. of work since each week it scans about 1 000 journals.

It is however crucial in the introduction of such a service to learn from the users' responses the value they set on the system's current awareness function. In the following pages we will therefore comment briefly on some results regarding current awareness and information use among the users of the SSCI.

1.3 Users' attitudes towards current awareness

The information needs and uses of scientists have been under investigation for the past 20 years, whilst those of social scientists have only recently been studied. Among the studies on social science information a project called INFROSS may be mentioned. This project was conducted as a broad study of the information needs and uses among English social scientists

(Line 1972). Another project has been conducted in the Soviet Union (Goldberg 1971). From these and other studies it is obvious that scientists and social scientists differ from each other in their preferences regarding information channels. [1] Social scientists rely more on monographs than scientists, who make more use of journals. This implies that social scientists rely more on older documents, and thus have less demand for current awareness. This conclusion is consistent with the general picture of the social scientist obtained in INFROSS:

> "...He is not highly motivated to find references, nor well equipped in the means of doing so. Library catalogues are little used. Most formal information comes through reading primary journals and press and publishers' notices of books. Much information is conveyed through informal channels of various kinds. Personal files are organized in highly personal ways." (Line 1972)

As a background to the evaluation study we present here some results to illuminate the information habits and attitudes of the user group.

In the survey, conducted after the test-period, we posed a question about the users' subjective estimation of their own current awareness. About 57 per cent considered that their (present) knowledge of current research was good or very good. When asked about the need for the rapid dissemination of journal articles about 57 per cent considered this important or very important. Another indication of a user's current awareness is his attitude towards cumulative research. The respondents were asked to consider the following statement:

> "If a scientist is not certain whether his problem has been studied or is being studied by others he should wait to decide on his own approach until he is certain that his research will fulfil a scientific need for knowledge."

1) See for example Garvey et al (1971)

The degree of agreement with this statement is shown in Table 1.1 below.

Table 1.1 Attitude towards cumulative research. Percentage distribution

Strongly agree	Agree	Undecided	Disagree	Strongly disagree	N
14	24	25	25	12	(317)

We found that 38 per cent clearly agreed with this statement concerning the necessity of cumulative research.

The conclusion we can draw from these results is that we cannot expect the social scientists to have more than a moderate desire for current awareness. This is also supported by the pre-test survey where the majority stated that once a week was too often to receive a new reference list.

There may be many explanations for these results. A large proportion of the user group are not PH.Ds, only a few are engaged in full-time research and many pursue applied research. Another factor which may be mentioned is the extensive use of technical reports as a predominant form of publication. These reports are often written in the author's own language, thereby limiting the possibilities of communication.

The use of the SSCI as a tool for information retrieval should also be considered in the light of existing information habits. From the point of view of the user, the introduction of the SSCI simply means that an additional information source is added to the sources already available.

1.4 The potential of SSCI for information retrieval

In the previous section it was noted that social scientists often lacked appropriate means for finding references in comparison with

those means available to scientists. To get some idea of the possible importance of the SSCI as a means of reference retrieval we here present some data about how the SSCI users generally retrieve references and what sources are used.

Table 1.2 Percentage distribution of source preferences. Most commonly used source.

Reference lists in books or journals	66
Bibliographies and abstracts	21
Ask another person	5
Card-catalogues at libraries	4
Search of book-shelves at libraries	3
Other sources	1
N	(387)

From Table 1.2 we can conclude that users have an outstanding preference for books and journal articles as sources for reference retrieval. Bibliographies and abstracts are the second most-used sources, bot are far less frequently used than books and journals. In short, primary sources are more often used than secondary sources.

From the survey we also concluded that only 40 per cent of the users had used Dissertation Abstracts and only 20 per cent knew of Current Contents in Social Sciences.

This marked lack of use of secondary sources for information retrieval should imply that for many of the users the SSCI ought to be a complementary source of great importance. This assumption

is also supported by the fact that 87 per cent of the users expected the SSCI to be of great or very great importance for their research.

Another aspect of the information habits of the users is their use of scientific journals. Since the SSCI is a secondary source for scientific journals it would be of interest to know how much of the SSCI coverage of journals the users cover themselves by reading journals. The potential value of SSCI for the users could be estimated by counting how much extra information, in terms of journals, SSCI can offer in relation to existing patterns of information use among the SSCI users.

In the pre-test survey we asked a subsample of users to name the five most commonly used journals. These journals were then compared with the SSCI coverage. It was found that this subsample together covered 127 of the journals covered by the SSCI, that is 13.2 percent of the SSCI coverage. This number is very small but should be interpretated with care since only five journals were mentioned. The potentially relevant journals that SSCI covers must however be assumed to be quite large in relation to user coverage.

SSCI can be seen as an interdisciplinary data-base affording users the opportunity of retrieving references from journals not directly concerned with their subject area. From the pre-test survey it was possible to study the degree of specialisation regarding reading of journals covered by the SSCI. Table 1.3 shows that very few journals in the SSCI coverage were read by more than 4 users.

Table 1.3 Number of users regulary reading SSCI covered journals.

Number of journals in the SSCI coverage:

Read by none	836
Read by 1 user	89
Read by 2 users	27
Read by 3 users	15
Read by 4 users	8
Read by 5 users	6

Since the user's own coverage of social science journals is so specialized as Table 1.3 indicates, the SSCI should be able to offer the possibility of both a broader and a more common use of journals.

The introduction of the SSCI seems to have raised high expectations among the users. It should be noted that these expectations are high despite only a moderate demand for current awareness. This may be assumed to influence the user's evaluation of the service.

These high expectations consequently resulted in a large number of participating users. The number of users, about 600, was however far too great for a completely personalized service, and the IDC could not deliver primary documents, which had to be accessed through the local university libraries. Furthermore, the documentalists at the IDC lacked experience from serving social scientists. The test period was limited to 13 weeks, which is too short a period for the system to reach its mature state. Bearing in mind these initial practical limitations, the evaluation study conducted here should not be regarded as a definite evaluation of the SSCI among Swedish social scientists. The evaluation is a study of an IR-system in its immature state. The aim of the evaluation is twofold. Firstly, the evaluation study is an attempt to find possible ways of improving the service. Secondly, it is an attempt to develop a general strategy for such evaluations.

II EVALUATING INFORMATION RETRIEVAL SYSTEMS

In short, evaluation of information retrieval systems is an analytical procedure having fundamentally three pruposes (Lancaster 1968):

(1) To determine how far the system is satisfying user requirements,
(2) to determine sources of system failure, and
(3) to determine how these failures may best be remedied.

In order to make such an evaluation we have to start by identifying those strategic phases in the IR-process that are crucial for its performance. An IR-system can be seen as a process that begins with the first contact between user and system and continues with repeated outputs and user evaluations. In each phase several aspects of evaluative interest can be studied. For each phase we may select goals, the achievement of which is important. Then these goals must be transformed into measurable variables or criteria. If these criteria are not met (to a greater or lesser degree) we can define instances where the system failed.

Having located system failures we must have some theory about why the system failed in order to find remedies. To obtain explanations of this kind we are bound to have some sort of theory containing assumptions that describe the relationship between the IR-process and factors affecting it. The principal demand as regards this theory is that the assumptions, at least to some extent, can be tested empirically in the evaluation study.

In this chapter we shall initially discuss some general problems involved in the evaluation of IR-systems. Then we shall describe different phases of the IR-process and select the appropriate goals with their corresponding criteria. At the same time assumptions about how these criteria or variables are affected by other variables will be presented.

2.1 Some general evaluation problems

During the last few years there has been a rapid increase in computer-based information retrieval systems. At the same time there is a growing interest for an evaluation of these systems. Evaluations of IR-systems have largely emphasized the study of relations between costs and benefits or effectiveness. Costs for retrieving references have been compared with the amount of retrieved information in terms of precision and recall. using different indexing principles such as titles, titles plus keywords and titles plus abstracs (Scheffler et al 1974). Usually the interest in economic evaluation is a consequence of the vast cost of computerized IR-service.

In these studies the main effort is devoted to evaluation of the end product of an IR-process in relation to costs. In other words: How costly is it to achieve a certain kind of system output? In our view this is only part of the whole matter. The goals of an IR-system contain aspects that are not primarily related to economics and system output. In order to get a good overall performance many subgoals must be achieved in different phases of the IR-process. Therefore the evaluation of IR-systems must also be centered on the evaluation of different phases of the process. Otherwise we will get no explanation of why the system output was of a particular kind.

Evaluation most commonly implies that you want to compare the result of some practical action with its stated goals. In order to make such an evaluation we must define goals, transform these goals into measurable criteria and decide to what extent these criteria should be met to fulfil the goal. So far the evaluation is merely descriptive. That is, we only arrive at conclusions regarding whether the stated goals were achieved or not. However, to be of practical value we must find factors that explain these described facts. The criteria can be seen as dependent variables which can

be affected by other variables, usually called independent variables. We must know why the goals were not achieved in order to find the remedies necessary to improve the system. In short, the evaluation problems of a general nature consist of defining goals, transforming these into measurable criteria and finding causes for the empirical findings regarding goal achievement. We shall briefly discuss these problems one by one.

At first glance, the problem of defining goals for an IR-system seems to be fairly easy: The aim is to satisfy user information needs. The corresponding criteria might be a question asked of the users: Were you satisfied or not? If all the users are satisfied we can also be satisfied. But if some users are not wer just have a descriptive fact of little value. We can hardly conclude that the system is of no value for these users because the criteria does not indicate where the causes for dissatisfaction are to be found. There might be a lot of causes; it may be that the users lacked the knowledge to give an adequate description of their special interest area, that is they gave a false definition of their information needs. The documentalists who constructed the search profiles may have misunderstood the meaning of user queries. The data-base didn't cover the kind of information that was asked for etc. These examples indicate that we must have a more specific definition of goals. For example: The users ought to have some degree of knowledge about their special interest are, the interaction frequency between user and documentalists must be so high that misunderstanding cannot occur, users with special kinds of interest areas that the data-base doesn't cover should not be contacted or served. In short, we need to define the goals related to different phases of the IR-process in order to locate system failures and to explain them.

To specify system goals in relation to specific phases of the process is not enough. We must also discuss to what extent it is reasonable to stress the necissity to fulfil these goals. Can we demand that

the IR-system performs at an optimal level in each phase? In some way we must define reasonable demands in relation to system resources in terms of the time, money and skill that is provided for running the service. In relation to time we can conclude that an IR-system generally has to be run over a long period before it reaches its full maturity or "steady state". If an IR-system is newly introduced, we must be aware that we are evaluating a "transition state" of the system and therefore we must adapt our demands. In the "transition state", when the system is being introduced, there are all the problems of implementation that must be solved. Implementation means i.a. that the system must be made practically available and this implies that there must be an adequeate definition of its users: The users' descriptions of their information needs, in the form of queries, must correspond to their needs and to the general way of defining research problems in the special interest areas. System resources must be satisfactory in terms of money and personnel. If these implementation problems are not overcome, we can hardly say in justification that we are evaluating an IR-system regardless of other aspects that can be of crucial importance.

When empirical measures of goal achievement are found we must look for factors that explain why the system performed at that level in that phase. In doing so we distinguish between factors or variables that are internal to the IR-process and factors or variables that are external. The distinction is important because it indicates whether the factors are controlled by the system (process variables) or whether they belong to the surrounding milieu and are only indirectly controllable. It is for example possible to control the way in which queries are formulated by informing users about how this is to be done. But it is hardly possible to control a user's knowledge of his special interest area, a factor that can be assumed to have great importance for the whole IR-process. On the basis of causes which affect system performance in the different phases it should in principle be possible to decide what actions to take in order to improve the

IR-process.

On the following pages we present an evaluation strategy that, at least to some extent, takes these general problems into consideration.

2.2 Phases and criteria of the computerized IR-process

A computerized IR-process generally involves the following phases:

- Introduction
- Query formulation
- Profiling
- Output and user evaluation
- Termination

For each phase there are questions of special interest to be answered. Did the IR-process achieve those goals that are important to achieve in each phase? We shall discuss what these goals might be, their corresponding criteria, and discuss factors affecting these criteria.

Introduction: The system is introduced to a population of potential users. The special qualities of the system are presented and users are asked to participate. Users that accept the invitation constitute the system's user population.

The general aim of an experimental IR-service is to estimate its ultimate value for a group of potential users. This is partly a matter of selecting an appropriate group of users to whom the service is introduced. In short, it is important to have an adequate definition of potential users. To evaluate this we must find out if there are other potential users who were not asked to participate, users who might have gained some value from using the service.

In order to arrive at an adequate definition of a potential user group we must have some idea about the structure of the scientific field that the IR-system is mainly directed towards. For example: In what organisations do we find persons who are engaged in social science research and will be interested in keeping up-to-date with current research in this field? One of our criteria will be as follows: Was the IR-system introduced to *all* the *relevant* persons? If there are potential users of the service who were never defined as such there is always the risk of testing the IR-system for only a limited group, thus narrowing our chances of estimating its ultimate value for social scientists in general.

An adequate definition of relevant potential users also implies that we know to what extent users aimed at using the service for instrumental purposes. If for example, the users' decision to participate is based on anomalous interests, his evaluation of the IR-service will not be based on the satisfaction of information needs in the instrumental sense. It is primarily the current awareness function of the IR-system that is to be evaluated and not its expressive functions in terms of a technical venture etc. If the users' motives for participating involve such aberrant interests then the definition of potential users is not adequate. Naturally, we want to have users that really intend to use the service in their daily work.

Another aspect or importance in the introductory phase is whether users possess enough knowledge of system functions to be able to make adequate queries or to initiate profile changes. It is important to establish some interaction competence so that users can utilize the system optimally. Often there is initially a great need for user feedback so that search profiles can be improved. When profiles are running well, the need for user-system interaction is less. What we want to know here, in terms of criteria, is the amount of knowledge about system functions that users possess (knowledge about search strategies, profile logic etc.)

Later on, when these questions are answered, we can look for factors affecting user knowledge about system functions. Experience of other IR-systems may be of importance and other experience based factors.

Query formulation: Users state their special area of interest for which information is needed. This is usually made in terms of relatively short verbal descriptions. Lists of references and periodicals that users consider to be of special interest are often enclosed.

It is of importance for the whole IR-process how well information needs are defined in the form of explicit queries. Firstly, the descriptions of a user's special interest area ought to contain the concepts or language commonly used by researchers in the area in order to ensure an adequate correspondence between the query and the international research literature. Secondly, when queries are translated into profiles it is important that the queries are so distinct semantically that the risks of misunderstanding are minimized. In this respect, some general demands regarding queries can be stated: They ought to contain a short and distinct description of the special interest area, a certain minimum of search terms. References and journal titles can also be enclosed in order to allow documentalists to study these as a complement to the information given in the queries.

Profiling: After receiving a user query a documentalist will construct a secondary representation of the query in the form of a search profile. Ideally, profiling also involves checking the query/profile correspondance by asking users.

Profiling can be defined as the operative activity needed to get a secondary representation of queries into profiles. This activity mainly consists of discovering relevant search terms and combining

these with a search logic. Profiling can be seen as an interpretative task. Consequently the general criteria for evaluating the profiling phase revolves around whether the correspondence between the meaning of queries and the profiles is adequate.

The amount of effort expended on profiling is determined by several factors. The degree of clarity and completeness of the queries if basic for the construction of adequate profiles. If the degree of clarity is low, documentatlists themselves often have to find a more specific definition of the users' interest areas and also to look up relevant search terms. Here the competence of the documentalists in the subject field of the query is a strategic factor. Resources in terms of money and time are also of great importance. The intermediary link between user and system that documentalists constitute can be excluded if users construct the profiles by themselves. This will, however, make advanced user education necessary.

Output and
user evaluation: Users receive printed lists of references. Users evaluate the references according to relevance. User evaluation of system output is a necessary condition for improvement of search profiles.

The output of an IR-system has traditionally been evaluated in terms of the amount of information retrieved (output), the degree of relevance in output (precision) and the degree to which relevant information in the data-base has been retrieved (recall). These criteria tell us the degree to which the system has succeeded in the selection of information.

System output is directly affected by two main factors. The search profile determines what information to select and the data-base what to select from. In terms of information processing the profile can be seen as a filter by means of which stimuli are selected.

The data-base with its references constitutes the stimuli that are to be filtered. The interplay between profile and data-base results in system output.

For the user the value of an IR-system is estimated in terms of how well it satisfies his information needs. It is the end product of the interaction with the IR-system that is of primary importance for the user. He is of course more interested in what he gets from using the system than how he gets it. The normal criteria used are precision and recall. In this study we supplemented these with user satisfaction, which gives us a more complete indication of the system's overall performance.

User evaluation can also be extended to include the courses of action the retrieved information initiates. What effect has the IR-service on the users' current awareness of research? Have users' information gathering habits changed as a result of the IR-service? Have users integrated the information received into their daily work?

User evaluation of system output can be determined by a large number of factors linked both to the user himself as a person, and to his kind of work. For example, users with high current awareness will probably demand more of information received than those users who know little about current research. Another factor is how well the data-base covers research in comparison with users' scientific reference groups of the scientific community. Only a few of these factors, external to the IR-process, but that may affect system performance will be studied in this evalution.

2.3 Summary of the evaluation model

In this chapter we outline a general model for evaluation of computerized IR-systems. The main feature of this model is that

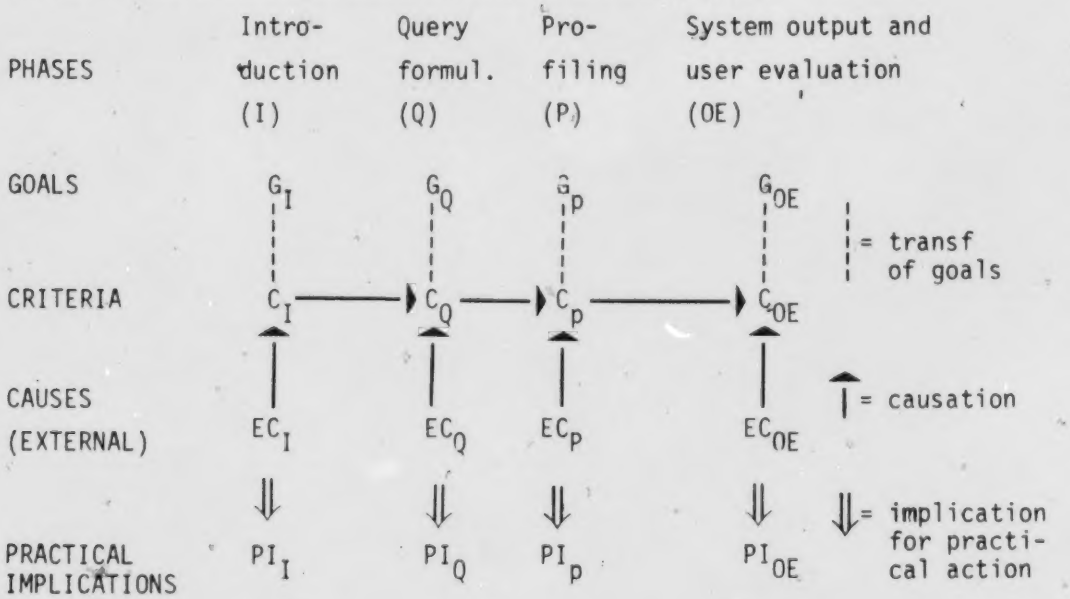
it allows us to treat the IR-system as a process. The aim of this model is to make it possible to identify causes behind system performance, not just to measure whether or not all went well.

In order to identify such causes the evaluation should be centered on different phases of the IR-process (Introduction - Query formulation - Profiling-System output and user evaluation - and Termination). In each phase there are some variables or criteria that are crucial for the IR-process. These criteria correspond to the specific goals of each phase. When the empirical results are compared with the goals we can determine the degree of goal achievement.

When this is done we can deal with three kinds of causes. Firstly, there are all the factors internal to and specific for each phase. Secondly, there are external factors, those not directly involved in the IR-process, that can affect the performance of each phase. Thirdly, the performance of each phase is affected by the performance of previous phases. Ideally, the evaluation study should take all three kinds of causes into consideration in order to determine their relative effects on the overall system performance.

The general evaluation model can be summarized by fig. 2.1.

Fig. 2.1 A general evaluation model for computerized IR-processes.



III DESIGN OF THE EVALUATION STUDY

In this chapter we present the design of the evaluation study and discuss the quality of the data gathered. The study mainly follows the evaluation strategy outlined in chapter II.

The desire to collect data from each of the different phases of the IR-process follows logically from the evaluation strategy. This makes it possible to describe the components involved in the process (users, queries, profiles, data-base). It is also a necessary condition if one wishes to analyse the interrelationship of different phases and estimate their impact on system performance.

3.1 Definition of the studied population of SSCI-users

The service which was offered to about 1 100 social scientists and some earlier customers of the IDC, elicited about 600 acceptances in total, not all of whom participated for the whole test period. In order to get a test group which was homogenous in this respect the population for the study was defined as those users who participated from the first week of the test period which 497 users did. This population was composed of 95 per cent social scientists at the universities.

The data gathering methods used are described in the following sections.

3.2 Pre-test survey - before the test period

During July and August 1973, before the beginning of the retrieval service, a survey was conducted based on a systematic sample of users. The aims of this survey were to make it possible to measure effects of SSCI-usage on users' information use and current awareness.

Another aim was to test various items and to get some knowledge about the user group and its expectations and demands regarding the service. At present, the survey has not been analysed as to the effects of SSCI-usage. The number of responses in the pre-test survey was 151, a response rate of 67 per cent.

3.3 Evaluation of reference lists - during the test period

During the whole test period each reference list was followed by an evaluation form. The users were asked to evaluate each reference according to a six-grade scale (See Appendix V). During the test period 26 766 references were evaluated. Since some users did not participate from the beginning of the test period their evaluations are excluded from the analysis. In total about 60 per cent of the references sent out were evaluated. It is hard to estimate in terms of representativity the precision rate for the non-evaluated references. The non-responses partly seem to be the result of exhaustion among users, which led to incomplete evaluation over the 13-week period. The number of users that made evaluations corresponds to a response rate of 74 per cent.

3.4 After the test period

Main survey

The main survey was sent to all users participating in the test. The overall response rate was about 67 per cent (355 out of 497 users answered the questionnaire). [1)]

The aim of the main questionnaire was not strictly limited to the evaluation of the SSCI. We also wanted some knowledge of the information needs and uses of social scientists, since such studies are very rare.

---

1) For comparison the large survey among English social scientists conducted by INFROSS yielded a response rate of 50 per cent.

Output

From search statistics the number of printed retrieved references for each user was noted (output).

Profile and query characterization

In order to obtain data regarding profile characteristics 182 profiles were systematically sampled. This subsample represents 37 per cent of the profiles run initially. Each profile was described in terms of various attributes.

A subsample of 72 queries was also analysed and described according to a few characteristics.

3.5 Comments on the data gathered

In total the number of different data-gathering methods used in the study amounts to 289 variables studied. In this report we have mainly used the survey, the evaluation of reference lists and query and profile characteristics.

Following the evaluation strategy we can summarize by stating that the evaluation should include estimations of system performance and analyses of its causes. In estimating system performance we can rely on two extensive data collections - the main survey and the interest evaluations (of reference lists). In the analysis of causes we rely most on the main survey supplement with profile and query characteristics from a subsample of users. In this section we discuss the effects of non-responses for the survey data (the interest evaluations are discussed briefly in section 3.3).

Firstly it should be stated that the main survey was conducted as a study covering the whole population of users who participated from the beginning of the test period. Hence, the problem of

representativity concerns only the impact of non-responses. However, there is another problem - does the user population represent adequately the various disciplines and Swedish social scientists in general? This is discussed in chapter IV. Here we concentrate on the effects of non-responses.

As indicated in section 3.4 the response rate in the main survey was 67 per cent. The question is whether the non-responses deviate in any respect that is important for the study.

It is crucial in any estimation of performance to know how far response rate depends on system performance. A negative correlation between response rate and system performance implies an underestimation of total system performance, while a positive correlation implies an overestimation of total system performance.

In considering the accuracy of our estimations of system performance some measure of sytem performance without response failures must be used. The only measure available from the search statistics at the IDC is output (the number of printed retrieved references) which was used to analyse the relationship between response rate and system performance (Table 3.1).

Table 3.1 The relationship between output and response rate in the main survey. Percentage distribution.

	OUTPUT		
	0-40	40-100	100-
Respondents	61	71	75
Non-respondents	39	29	25
N	(251)	(124)	(122)

From Table 3.1 we can see that the response rate is positively correlated with system output (a correlation analysis using "gamma" yields a coefficient of -+ 0.23, see Goodman and Kruskal 1954). However, output only partly measures system performance. A lack of system output does not necessarily mean that users are dissatisfied. In some cases a computerized current awareness service is used as an instrument to check users' own references gathered previously, where a lack of output confirms that users' own reference gathering is satisfactory.

Considering the effects of non-responses on the estimations of performance we can conclude that system performance is somewhat overestimated, but the differences are very small and do not change the general picture.

3.6 Comments on the methods used in relation to the evaluation strategy

The evaluation strategy outlined has implications for data gathering and analysis as well as for the administration of the service. The evaluation strategy implies that, ideally, goals are made explicit for each phase in advance, data cover each phase of the IR-process completely and theories can be used to explain the outcomes in each phase etc. Of course, this must remain the long term goal of this kind of research. To complete all the analyses, that the evaluation strategy implies are desirable, was not possible in this study and would have demanded far more resources for data gathering and analysis. The evaluation strategy outlined here has, as far as we know, not been systematically used earlier. This evaluation study should thus be regarded as an early contribution to the testing and development of an evaluation strategy. However, two limitations of this evaluation study in realtions to the evaluation strategy will be discussed.

Firstly, goals were not made explicit for each phase before the test period. If goals had been made explicit at the beginning of the test period, this would have made it possible to concentrate the evaluation of performance in each phase and make continual improvements during the test period. In this case, this was not possible since the evaluation study was initiated *after* the introductory phase. The test period was intended to be a trial-and-error experiment due to the fact that very little was known about social scientists as a user group. Both the data-base used and social science as a field for information services were new to the IDC. Furthermore, information needs and users in the social sciences are generally investigated to a small extent.

Secondly, data gathering was not complete in all phases. The resources for gathering data regarding the query formulation and profiling phases did not allow for more than a partial coverage of these phases. The effect of this in relation to the evaluation strategy is that the possibilities of using complex methods for data analysis are restricted.

IV INTRODUCTION OF THE SSCI

The first problem that the IDC was confronted with was to introduce the service to a group of potential users, in order to obtain an adequate test group for the evaluation study. The IDC decided to offer free service during thirteen weeks to all researchers and teachers in social sciences at six Swedish universities since these tapes were offered free of charge from ISI. About 1 150 social scientists received the offer by personal letters. Of these about 500 agreed to participate and they also submitted descriptions of their special interest areas, on which profiles were then constructed. After profile construction the IDC personnel held working seminars at the universities in order to establish closer contact with the users and to discuss profile constructions. These seminars were held in May 1973.

When users were offered the service they were told that they had to supply a continuous evaluation of the references that were supplied. They also had to answer a questionnaire at the beginning and at the end of the 13-week period. In this chapter we discuss some problems regarding the introductory phase that may be regarded as important for the whole SDI-process.

4.1 The user group

The general purpose of the SSCI-project was to determine the value of a computer-based current awareness service for Swedish social scientists. Before the service was offered it was concluded that the majority of social scientists whose activities imply a need for current awareness of scientific information are located in the universities. There are of course some research institutes with qualified social research that could have been considered. However, to reach these groups would have made a more detailed survey necessary. Thus, the time and money available dictated that the universities should be chosen as the first test group. Later this

group could be extended to special institutes and practitioners.

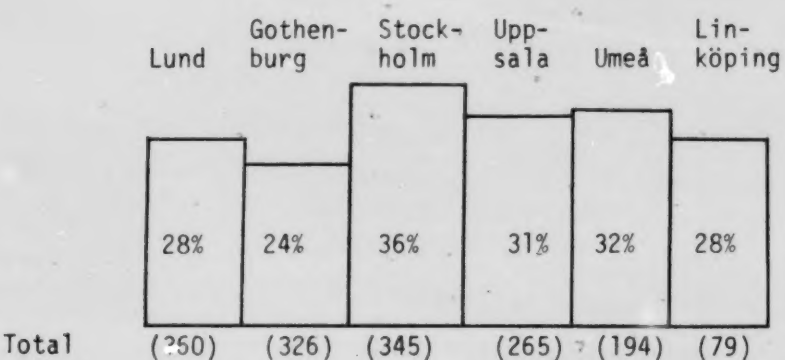
User studies have shown marked differences in preferences for information sources among researchers, teachers, managers and practitioners. Researchers and teachers rely heavily on printed sources such as periodicals and textbooks. Practitioners and managers rely more heavily on informal and semi-formal sources (unpublished material, trade magazines etc.) (Saltzman 1968, Allen 1966, Menzel 1970, Rosenblom et al 1967, American Psychological Association 1969).

The SSCI can be seen as a formal source covering periodicals and therefore researchers and teachers should gain most benefit from it. According to this reasoning the potential user population was defined as:

> "all researchers and teachers of social science in Swedish universities (i.e.) in Umeå, Uppsala, Stockholm, Lund, Gothenburg and Linköping".

This population received an offer of the service, free of charge by personal letter. The acceptance rate was fairly similar at all universities.

Fig. 4.1 Percentage of users at the six universities.



We can conclude that the SSCI-users are fairly equally distributed among the six universities. However, not all those listed in the university catalogues received the offer (some of the catalogues used were not complete). The acceptance rate therefore is higher than Fig. 4.1 indicates. In total, 1 163 individuals received the offer and from the six universities 473 individuals accepted. This yields an acceptance rate of 41 per cent (24 non-academic users are excluded).

Furthermore, if we look at the distribution of SSCI-users according to discipline (Table 4.1) we can see that there is a good representation of users from each discipline. The only discipline that is somewhat underrepresented is political economy. Political economists represent 4.2 per cent of the SSCI-users and 11.7 per cent of the potential users who received the offer.

Table 4.1 Distribution of SSCI-users by discipline and distribution of potential users by discipline.

	SSCI-users		Potential users/Offered	
	Number	Per cent	Number	Per cent
Business Administr	68	13.7	114	9.8
Political Economy	20	4.0	135	11.6
Sociology	74	14.9	173	14.9
Psychology	68	13.7	149	12.9
Pedagogics	52	10.5	135	11.6
Economic History	14	2.8	44	3.8
Cultural Geography	37	7.4	98	8.5
Political Science	28	5.6	93	8.0
Other	105	21.1	218	18.9
Outside the university	24	4.8		
Unknown	7	1.5		
	497	100.0	1.159	100.0

Looking at the two percentage distributions in Table 4.1, we can conclude that the distribution of SSCI-users according to discipline resembles the corresponding distribution of potential users.

Another aspect of user population that is relevant for the discussion about the SSCI-users as an adequate test group is to what extent the users were engaged in research activity. The SSCI mainly contains information that is relevant for researchers. Therefore, the SSCI

should mainly be tested on social scientists activitely engaged in research. As we can see from Table 4.2, most of the users had at least some research activity in progress - about 90 per cent of the SSCI-users.

Table 4.2 Distribution of research activity among SSCI-users.

Research activity	Number of users	Percent
- 0 h./week	28	8.7
1 - 15 h./week	106	32.8
16 - 30 h./week	101	31.3
31 - h./week	88	27.2
	323	100.0

Finally, at the beginning of the 13-week period the SSCI-users also contained persons outside the university (about 5 per cent of the SSCI-users). These users were mainly employed in industrial or government research institutes or investigation departments.

Looking at the SSCI-users as regards their relevance as a test group we can say that the user population represents a good coverage of Swedish social scientists as to geographical location. discipline and research activity.

4.2 Information about the aims and functions of the SSCI-service

A computer-based current awareness service has to be seen as a radically new way of retrieving scientific information, especially for social scientists in Sweden. The innovative character of the

SSCI is shown simply by the fact that only 15 per cent of the users had come into contact with any computer-based information service. This implies that users initially knew little about the aims and functions of the SSCI-service. This put a heavy obligation on the managers of the system to inform the users about the concept of a computerized information service. During the implementation of the system it is important to have close contact with users in order to inform them about how they should change their profiles if their searches achieve poor results, how to retrieve original documents etc. If users do not possess this competence there is always a risk that they will not initiate profile changes if so needed and will not therefore make full use of the possibilities of the system.

As noted earlier the IDC-personnel held working seminars at the universities. The participation rates at these seminars at the different universities is shown in Table 4.3.

Table 4.3 Participation rates at the working seminars held at the six universities.

	Percent participators	Total number of users
Lund	15	102
Gothenburg	51	79
Stockholm	40	124
Uppsala	24	83
Uneå	40	63
Linköping	68	22
		473

In general, only 35 per cent of the user population participated in the working seminars. To evaluate the effect of participation in working seminars on interaction competence in dealing with the system Table 4.4 should be studied.

Table 4.4 Three indicators of user knowledge of system functions by participation in working seminars. Percentage distribution.

a) "Do you feel that you have got enough information about the SDI-service to understand its function?"

Answer	Participators	Non-participators
Yes	88	74
No	12	26
	(76)	(255)

b) "Have you by yourself proposed any changes in your profile so as to get better lists of references?"

Answer	Participators	Non-participators
Yes	47	32
No	53	68
	(77)	(255)

c) "Do you consider you have enough knowledge about the various search possibilities that exist in order to suggest how your profile should be changed?"

Answer	Participators	Non-participators
Yes	61	44
No	39	56
	(72)	(252)

From Table 4.4, it is obvious that participation in working seminars generally had a positive effect on user knowledge of system functions. Another result is also interesting - only about 50 per cent of the users knew how to change their profiles (see item c) in Table 4.4). There is obviously a great lack of knowledge among users which may have a negative effect on their disposition to initiate profile changes if necessary.

V QUERY FORMULATION

When users accepted the invitation to participate they also submitted a short description of their special interest area. Profiles were then constructed on the basis of these descriptions. To form a basis for profile construction queries ought in principle to fulfil the following criteria:

- Clarity: A necessary condition is that the description of interest area is clear enough to be comprehensible. In order to be comprehensible a query ought to contain:
  a) at least some words which can be utilized at search terms and
  b) at least one sentence describing the logical relationship between these words.
  c) a list of relevant references and
  d) a fuller description of the interest area.

- Correspondence with data-base: There must be at least some correspondence between the interest area described and the content of the data-base. This does not only imply a correspondence between interest area and data-base but also between the types of documents asked for and those available.

These criteria must be fulfilled to some degree if there is to be at least a theoretical possibility of constructing an adequate profile representation of queries. In practice, because of the limited resources that are available, a more than minimal fulfillment of the criteria is necessary if service demand and available resources are to be balanced. In the case of this SDI-service, where the number of received queries was extremely high and resources for profile construction quite limited, the importance of query formulation is emphasized.

5.1 Degree of clarity in queries

As mentioned above, clarity can be described in terms of the number of search term and sentences in queries. From a sub-sample of user queries we obtained the following distribution (Table 5.1). Nearly 50 per cent of user queries did not contain a single search term. This implies that the documentatlists had to discover search terms terms themselves using term frequency lists, thesauri and other aids for profile construction. Considering the vast number of queries that had to be dealt with in this way, it is obvious that a great burden was placed on the IDC.

Table 5.1 Percentage distribution of search-terms and number of sentences in user queries.

Number of search terms	Percentage	Number of sentences	Percentage
None	49	None	14
1-2	10	1	32
3-4	12	2	29
5-6	7	3	18
7-9	22	4-	7
	100		100
	(72)		(72)

Furthermore, the number of sentences in the queries was generally low. About 75 per cent contained between none and two sentences. Sentences in queries are important in describing the logical relationship between search terms. If the number of sentences is limited the documentalists have to make their own personal interpretations which easily will result in deviations from the intended meaning of a user.

The general lack of clarity among the queries can be explained by a number of factors. When the service was offered users were

encouraged to describe their interest areas in a few sentences. But this was done without any further indication of how the queries would be treated at the IDC. More complete queries would probably have resulted if the users had known the principles of profiling. Considering that the IDC usually serves customers in science and technology the necessity for query clarity is even more important. Another factor that may explain the lack of clarity is that only about 15 per cent of the users had used a computerized information service before.

The fact that the SSCI only permits title searches makes it necessary to use a greater number of search terms than is necessary with other search methods. This is emphasized even further by the somewhat loose terminology of social science which implies that a great variety of concepts are used to describe similar phenomena. Many search-terms are thus necessary to compensate for the lack of consistency in terminology and title labeling.

5.2 Extra information in queries

Profile construction can be considerably facilitated if the queries are supplemented by extra information, for example lists of references and fuller descriptions of interest areas. Only about 6 per cent of the users had enclosed lists of references and 6 per cent also gave a fuller description of their interest areas.

5.3 The correspondence of queries and data-base.

Part of the evaluation aim was to see if the data-base coverage corresponded to the information needs of the users in terms of their subject fields. In studying the correspondence between queries and data-base the operative activities of the IDC are excluded. Queries with a low correspondence to the data-base have, of course, little chance of gaining anything from the service.

If the correspondence is close it implies that the user population was well-chosen.

In order to get some measure of correspondence we classified a sub-sample of user queries according to the broad subject areas which are used in the list of journals covered in SSCI. Comparing the number of queries with the number of journals in the corresponding subject area we obtained the following result (Table 5.2).

Table 5.2 Correspondence between user queries and number of journals covered in query subject area.

Number of journals covered in subject area	Number of queries (cumulative percentage distribution)	
- 10	5.6	
11 - 20	9.8	
21 - 30	12.6	
31 - 50	30.8	
51 - 65	47.5	
66 - 95	47.5	
96 -	100.0	N=72

About 50 per cent of the queries studied correspond to subject areas with more than 95 journals covered in SSCI. This indicates that the data-base has an information capacity of relevance for the user group. The measure used above is however rather crude and results in an overestimation of the number of relevant journals for each query since most queries concern a limited part of the broad subject areas.

Another measure of correspondence between users' subject areas and the data-base is to compare the subject areas in the data-base with user disciplines. When this is done the number of potentially relevant journals for each discipline is rather large due to the

correspondence between users' disciplines and journal coverage in the data-base. From Table 5.3 we can see that the most frequent user disciplines are matched with the frequency of journals by subject area.

Table 5.3 Most frequent subject area among users and number of journals in the corresponding subject area of the data-base.

Subject area/ discipline	Number of users	Number of journals in SSCI corresponding to subject area
Sociology	74	89
Psychology	68	145
Pedagogics	52	84
Business Administration	68	50
Political Economy	20	78
	282	446

The five subject areas in Table 5.3 represent about 60 per cent of the users and correspond to about half of the journals covered by the SSCI. The remaining SSCI-journals cover about 40 other subject areas.

Some practical principles regarding query formulation can be mentioned here. In order to achieve greater clarity in queries, users should be better informed about the principles of profiling so that the need for clarity in queries is stressed. The need for extra information could be satisfied in many ways but the best way is to establish close contact between users and documentalists. The importance of such interaction was shown by one striking fact from the evaluation study. From the survey we can conclude that about 50 per cent of the users felt that the IDC personnel had too little information about their interest areas to construct adequate search profiles.

VI PROFILING

At the beginning of the SSCI-service the documentalists had to construct about 500 search profiles. Resources for profiling were limited, as has already been mentioned, and the activity also involved discussions with users participating in working seminars.

The profiling phase can be seen in two ways. Firstly, profiling can be regarded as the amount of activity or effort needed to construct profiles on the basis of queries. Secondly, profiling can be evaluated in terms of correspondence with queries. This correspondence mainly involves the translation of query meaning into profiles.

In this chapter we evaluate as far as possible these aspects of profiling starting with a description of profiling activity.

6.1 Profiling activity

Profiling acitivity can be defined as the activity expended on profile construction and as profiling modification. At the most general level profiling activity can be measured in terms of the number of profile versions that were made.

Table 6.1 Percentage distribution of number of profile versions.

Number of profile versions	Percentage of profiles
1	27.5
2	28.0
3	20.3
4	13.2
5 or more	11.0
	100.0
	(182)

Of the 182 profiles studied about 70 per cent were modified (about 45 per cent were modified three or more times). This means that documentalists had to modify nearly every profile at least once.

The amount of profiling activity can be further specified by studying the kind of modifications made (Table 6.2).

Table 6.2 The frequency of different types of profile modifications (first profile version).

Type of modification	Frequency	Percentage of profiles modified
Added search terms	970	54
Excluded search terms	497	32
Added groups	89	21
Excluded groups	57	10
Added search conditions	198	34
Excluded search conditons	186	24
Added hit conditions	11.238	60
Excluded hit conditions	6.496	40
Change of weights	83	7
Rearranged search terms	105	11
Change of terms/truncations	200	20

Total number of studied profiles: 182

The most frequent modifications were the addition of hit conditions (60 per cent of the profiles were modified in this manner), the addition of search terms (54 per cent) and the exclusion of hit conditions (40 per cent). Exclusion of groups, rearrangement of search terms and change of weights occurred more seldom. In reading Table 6.2 it should be noted that a change in search conditions may also result in a change in hit conditions etc. Since the modifications are not independent of each other, Table

6.2 gives only a rough estimation of the modification made.

6.2 Query - profile correspondence

The main aspect of profiling that is of evaluative interest is how well documentalists succeeded in translating queries into profiles in terms of meaning. Such correspondence is what we could call a highly qualitative matter, hard to measure directly. To do so, we would have had to make a special investigation, analysing and comparing the content of profiles and queries. An easier way was to include a question in the main survey and thus obtain the user's idea of how well his profile corresponded to his query. In an earlier study query-profile match has been estimated by comparing the number of positive and negative concepts in query/profile, seeing whether concept expansion in profiles broadens or narrows the field covered by queries and whether the search logic parallels query logic (UKCIS, 1974)

Table 6.3 Question: »Do you consider that the documentalists at the IDC received enough information about your special interest area in order to be able to make an adequate representation of your search profile?»

Answer:	
Yes	44.6 per cent
No	55.4 per cent
	(316)

Even though nearly every profile was modifed, 55 per cent of the users felt that their profiles failed in some way to correspond to their queries. The main conclusion from these results is that the need for profiling is much greater than the figures in Table 6.1 might indicate. This may be only a question of more resources for profiling, but all these figures obviously demonstrate the need for documentalists with a high degree of compentence in social science.

VII SYSTEM OUTPUT AND USER EVALUATION

The criteria most often used to measure the performance of an IR-system are output, precision and recall. These measures are defined in the following way:

Output = the number of retrieved references
Precision = the number of retrieved relevant references/output
Recall = the number of retrieved relevant references/number of relevant references in the data-base

Recall is not estimated here because of the great difficulty involved in estimating the number of relevant references in the data-base. Furthermore, measuring recall is not relevant when one is dealing with a continually changing data-base such as SSCI. The profiles are constants and the data-base a variable which implies that the number of relevant references in the data-base will always vary from search to search (Hjerppe, 1974). The users can also to some ext[illegible] usually at the change of precision, determine their level [illegible]ll.

The performance measures or criteria can be seen as the end product of the IR-process that has to be explained in the evaluation study. In terms of cause and effect, performance criteria constitute the dependent variables or effect variables. As mentioned earlier, the causes are those variables located to the previous phases of the IR-process, either internally or externally related to the IR-system. A test of how these factors effect performance is presented in the next chapter. In this chapter we only describe system performance leaving aside the problem of causality.

7.1 System output

Table 7.1 shows for the 13-week period the number of profiles that received output, the number of total retrieved references and mean references per profile, for every week.

Table 7.1 Number of profiles with output, number of retrieved references, and mean output per profile for 13 weeks.

Week	Number of profiles with output	Number of references on the tape	Total number retrieved references	Utilization percentage	Mean output per profile with output
1	370	1293	3.771	292	10.20
2	357	1090	3.338	306	9.35
3	362	1187	3.401	286	9.40
4	315	524	1.514	289	3.81
5	284	513	1.231	240	4.33
6	361	1130	3.262	289	10.14
7	352	1162	3.005	258	8.54
8	393	1654	5.719	346	14.55
9	402	2039	5.050	248	12.56
10	298	915	1.806	197	6.16
11	377	1368	3.301	241	8.76
12	364	1331	3.310	249	9.09
13	289	1656	4.499	272	15.57

Total number of retrieved references = 43.207
Mean number of profiles with output/week = 348
Mean number of retrieved references/week = 3.324
Mean number of output per profile/week = 7.04 (base 472)
6.69 (base 497), including those profiles that never received output)

Looking at the number of retrieved references during the period we can note fairly large fluctuations. This fluctuation depends to a large extent on data-base periodicity. This can be seen in the utilization percentage which varies much less.

The mean number of output per profile was about 7 references per week despite the fact that five per cent of the profiles never received any output.

We can go further in studying system output by looking at the distribution of output for all profiles over the whole period.

Table 7.2 Distribution of output for all initially run profiles over the whole 13-week period.

Output	Number of profiles	Percentage
None	25	5
1-40	226	45
40-100	124	25
100-	122	25
	497	100

We can see that 50 per cent of the profiles had an output of 40 references or fewer, and 25 per cent an output of 100 references or more. Five per cent of the users never received output. In terms of user satisfaction lack of output does not necessarily imply dissatisfaction since the data-base can be used as a checking instrument (see chapter III, page 20).

Another aspect of interest as regard system output is its regularity over a period of time. In Table 7.3 we can see that 34 per cent of the users received output every week, 27 per cent 10-12 times, 14 per cent 7-9 times, 10 per cent 4-6 times, 10 per cent 1-3 times and 5 per cent never.

Table 7.3 Distribution of profiles by number of received reference lists (output).

Number of received reference lists	Number of profiles	Percentage
None	25	5
1- 3	48	10
4- 6	52	10
7- 9	68	14
10-12	134	27
13	170	34
	497	100

The measure used in Table 7.3 can also be said to indicate fluctuations in data-base coverage. About 30 per cent of the profiles received output every other week. It is of importance to note the fluctuations in the data-base in order to determine the necessary number of searches that has to be run in order to test the relevance of the data-base for special interest areas.

7.2 User evaluation

During the 13-week period the users had to continuously evaluate the received references according to a six-category scale. About 60 per cent of the references retrieved were evaluated in this manner (Table 7.4).

Table 7.4 User evaluation of received references.

	Number of references	Percentage
Of immediate interest	4 917	19.8
Of interest, but already known	709	2.9
Of secondary interest	7 504	30.2
Undecided, since citation is not clear enough	1 564	6.2
Not of interest, in connection with my described interest area. Irrelevant.	10 046	40.4
Of no interest, as my interest area has changed since I described it to you.	110	0.4
	24 850	100.0

This user evaluation can be interpreted with varying rigidity. We can say that out of 24 850 references there were 19.8 per cent that were relevant. If we add the number of references already read we can say that the precision rate was 22.7 per cent. If references of secondary interest are also added precision is 52.9 per cent. This result can be compared with the results from user evaluation of other data-bases at the IDC. During the period 1967-1969 the corresponding precision percentages were 25.1, 31.9 and 62.4 respectively for a total number of 36 000 references evaluated.

The SSCI result can be said to be somewhat worse, but not much. This is a fairly positive result in that it relates to the "transition state" of the SSCI. However, a strong reservation must be made here - we really do not know the degree of relevance for the unevaluated references which totalled about 40 per cent. A possible hypothesis is in fact that the precision rate should be somewhat lower for these references; users that did not evaluate their references were probably more dissatisfied and thus less willing to spend their time evaluating reference lists.

In the main survey we posed some questions regarding users' overall evaluations of the received references. Table 7.5 shows users' degree of satisfaction in relation to the number of received relevant references.

Without doubt about 25 per cent of the users were satisfied with the number of received relevant references and about 50 per cent were rather dissatisfied. If this is a true picture, about 50 per cent

Table 7.5 Degree of satisfaction with number of relevant references.

	Percentage
Far too great	2.7
Sufficient	22.6
Somewhat too few	23.6
Far too few	41.2
None of any interest	9.8
	100.0
	(328)

of the users should feel that the SSCI gave them little or nothing of value in general. Table 7.6 shows the overall satisfaction with the SSCI-participation period.

Table 7.6 Users' overall satisfaction with the SSCI.

	Percentage
Very useful	3.9
Useful	27.2
Fairly useful	33.2
Of little value	27.5
Completely useless	8.1
	100.0
	(334)

The overall satisfaction is not as bad as Table 7.5 might imply (50% clearly dissatisfied users). About 65 per cent felt that they had had some benefit from using the SSCI and only 8 per cent were clearly negative.

Generally, the evaluation of the performance of an IR-system stops at the level of relevance evaluation made by users. The evaluation can be developed to include the courses of action that the user embarked on having received the references. Hence, evaluating the delivered information in its research context. What implications in terms of further action did the delivered references have for the users? Table 7 7 shows the different actions that were taken as regards the delivered references that were to be of interest.

Table 7.7 Percentage of relevant references that were retrieved in original form, read and utilized in work.

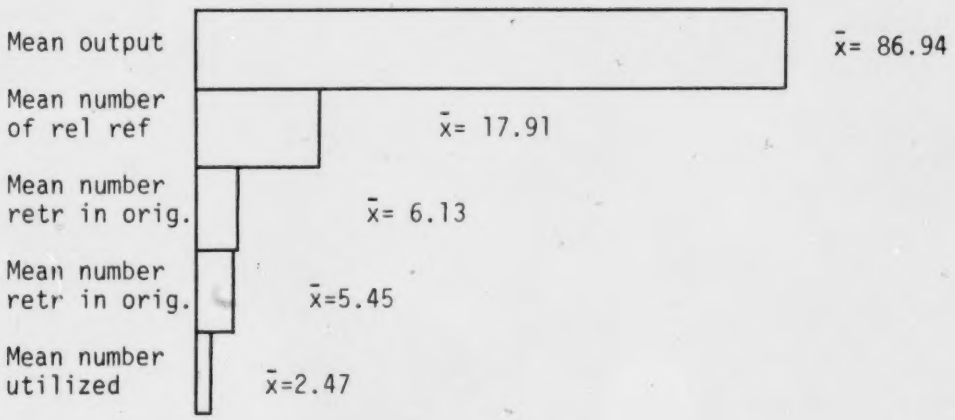
	Number of references	Percentage
Originally evaluated as relevant	4 337	100
Retrieved in original form	1 918	45
Read	1 670	39
Utilized in work	773	18

Data based on 313 user estimates.

From Table 7.7 we can see that about 45 per cent of the relevant references were retrieved in their original form on the user's own initiative, 39 per cent were read and 18 utilized in the users' daily work. If these data give a true representation (76 users did not estimate their output in this way) only about half the relevant references were so interesting at the time that users considered them worth looking for. This result should be interpreted with some reservation. Users may very well have considered those relevant references that were not immediately followed up as documents of future interest, i.e. they may be looked for when directly needed.

The IR-process can be described as a stepwise selection of information. Using this approach we can summarize the IR-process by estimating the mean number of selected references in each step (Fig. 7.1).

Fig. 7.1 The stepwise selection of references in the IR-process.



In the first selection step 80 per cent of the retrieved references were sorted out as irrelevant. In the next step only a third of the relevant references were retrieved in their original form. This indicates a restrictive attitude towards immediate retrieval of original documents. However, as to information availability about 50 per cent of the users had some difficulties in retrieving original documents. (See Table 7.8).

Table 7.8 Question: »Have you succeeded in searching for original documents?»

	Percentage
Always	15
Nearly always	35
Sometimes	27
Seldom	19
Never	4
	(265)

VIII FACTORS AFFECTING SYSTEM PERFORMANCE

The performance of the IR-system can be regarded as one final measure of the selective dissemination process. In the previous chapters we have described the different phases of this process in terms of system goals, independent of each other. Each phase will influence the following phases. For example, we suppose that the degree of user knowledge about IR-functions that is established in the introductory phase will affect the way in which queries are formulated as well as the user's propensity to initiate profile modification. The degree of clarity in queries may affect the degree of the corresponding meaning of profiles. Profile characteristics may, in a consistent manner, affect system output. The IR-process can thus be seen as a causal chain consisting of interrelated phases.

To improve an IR-process it is desirable to get some idea about how these causes affect the performance of the IR-system. We can easily argue that causation is a self-evident aspect of the IR-process, but it is far more difficult to determine the relative importance of various causes and a knowledge of this is, of course, essential for system improvement.

In this chapter we try to answer two basic qustions. Firstly, what is the relationship between system performance and various criteria located to earlier phases of the process? Secondly, what are the interrelationships between different phases and criteria of the IR-process?

8.1 Selecting a measure for system performance

In the previous chapter different measures of performance were presented. The measures used, however, measure different aspects of performance and one of the problems of finding measures of performance is thus that no *single* measure gives a complete picture of how the system succeeds in reaching its ultimate goal of satisfying

user information needs. For example, output does not measure the degree of satisfaction, it merely indicates the amount of information that users have to scan in order to find out if there are any relevant references. Precision is a measure of the capacity of profiles for sorting out irrelevant information. Combining output and precision gives us the number of relevant references that users received. Even if the amount of relevant references seems to be a reasonable measure of system performance it is obviously incomplete. A given number of relevant references might be too many for one user, too few for another and just enough for a third. The degree of satisfaction with a given number of relevant references has an immediate relation to the special qualities of information needs.

Taking these problems of measuring system performance into consideration we find it necessary to use a measure of user satisfaction which directly indicates the users' own estimation of system usefulness. This is hardly a difficult problem, we just ask the users to what extent they considered the SSCi to be useful for them. This measure has pronounced relations to output, precision and number of relevant references (Table 2.1). The users' estimation of SSCI usefulness can thus be said to partly cover these other aspects of system performance.

Table 8.1 SSCI usefulness in relation to output, precision and number of relevant references.

	Mean output	Mean precision	Number of relevant references
SSCI usefulness:			
Low	72.0	19.1	4.3
Medium	101.9	25.0	12.4
High	102.1	35.5	23.9

The users' estimation of SSCI usefulness probably also includes other aspects in addition to precision, output and number of relevant references. It is an overall measure of system performance, primarily measuring the SSCI's capacity to fulfil user requirements of a complex nature.

8.2 Introduction phase

One of the main purposes with the introduction phase is to establish some user knowledge about system functions in order to facilitate the interaction between user and system. In other words it is desirable to achieve a certain level of user competence in interacting with the IR-system. Interaction competence can be measured in many ways as noted in chapter IV and some indicators are shown below (Table 8.2).

Table 8.2 Usefulness of the SSCI in relation to some indicators of user interaction competence.

	Usefulness of the SSCI:		
	Low	Medium	High
Participation in working seminars (percentage of participators)	19% (134)	21% (115)	26% (111)
Percent feeling lack of knowledge about IR-functions	28% (134)	23% (114)	17% (111)
Percent feeling they have enough knowledge to initiate profile changes	44% (132)	46% (112)	52% (109)

If we compare users with low satisfaction and those with high we can conclude that users with low satisfaction generally had a lower participation rate in working seminars, felt a greater lack of knowledge about system functions and about initiating profile changes. They also felt that better contact with the IDC was more desirable.

From these results it should be obvious that the interaction competence possessed by users is related polsitively to satisfaction. This can be interpreted in many ways. Knowledge of sytem functions could affect satisfaction in terms of better adjustment of aspirations; if you understand the conditions and limitations of an IR-system you will adjust your evaluation of system performance in relation to these known

facts. Limited knowledge of system performance and functions may lead to inflated aspirations thus affecting the evaluation negatively.

In terms of interaction with the IR-system, limited knowledge may inhibit willingness to initiate necessary profile changes thus affecting system performance negatively and consequently also satisfaction.

In terms of system improvement there seems to be a need for greater effort at improving user interaction competence. This could be achieved mainly by closer contact between users and documentalists, information campaigns etc.

8.3 Query formulation

In chapter IV we discussed various aspects of queries that can be assumed to be of importance for the IR-process. Good queries were those containing a certain minimum of search terms and sentences, with extra information enclosed, and corresponding well with the data-base coverage. Table 8.3 shows the relationship between satisfaction and these aspects of query clarity and query correspondence with data-base.

Table 8.3 Usefulness of the SSCI in relation to query clarity and query correspondence with data-base.

	Usefulness of the SSCI:		
	Low	Medium	High
Mean number of search terms	1.7	3.5	3.4
Mean number of sentences	1.9	1.7	1.8
Mean number of journals covered by SSCI within users' subject area	55.8	64.8	80.6
Number of queries	(9)	(49)	(39)

Comparing once again users with low and users with high satisfaction we can see that dissatisfied users had fewer search terms in their

queries and a lower correspondence with the data-base in terms of number of journals covered by the SSCI in their subject area.

It seems obvious that the number of search terms included in user queries is related positively as one indicator of query clarity to user satisfaction. The number of search terms is of importance since it can be said to determine the degree of exactness in queries. If the number of search terms is too low, documentalists themselves have to find new ones which may eventually lead to low correspondence between query and profile. In terms of practical action users ought to be encouraged to give more detailed descriptions of queries containing a greater number of search terms.

From Table 8.3 we also can conclude that the number of journals covered in SSCI in relation to user subject area is positively related to system performance: the probability of getting satisfactory service from SSCI is greater if the SSCI covers a greater number of journals in the interest area.

Summing up the results presented above we find that the criteria used for evaluation of the query formulation phase seems to have a marked relationship to system performance. In terms of practical action for improving the service it seems obvious that the query formulation has to be improved according to the principles outlined in section 5.3.

8.4 Profiling

Basically, a profile is constructed of search terms combined with a search logic that determines conditions under which a reference should be retrieved (See Appendix III). Generally, serach terms are grouped together and then the search logic combines groups of serach terms. The conditons under which references could be retrieved are of two kinds. 'Search conditons' refers to logical combinations of groups and 'hit conditions' to combinations of search terms.

Profile and data-base characteristics determine system output since these are the factors directly involved in the mechanized retrieval phase. Consequently profile characteristics should be related to the various performance measures at hand. Below four types of profile characteristics are related to the four measures of system performance used earlier (Table 8.4).

Table 8.4 Some profile characteristics in relation to measures of system performance.

System performance	Mean number of: Search terms	Groups	Search conditions	Hit conditions
Output:				N
Low	23.3	4.9	4.2	151.6 (88)
High	30.4	6.3	5.3	209.4 (79)
Precision:				
Low	27.5	5.8	4.9	224.8 (52)
Medium	31.9	6.1	5.7	225.8 (44)
High	23.9	5.1	4.0	121.2 (64)
Number of relevant references:				
Low	23.3	5.0	4.2	119.1 (52)
Medium	26.1	5.6	4.5	147.8 (36)
High	28.0	5.2	4.1	170.4 (47)
Usefulness of the SSCI:				
Low	24.7	5.6	4.3	115.7 (52)
Medium	26.7	5.5	4.3	216.9 (46)
High	27.2	5.1	4.7	138.0 (52)

The profile characteristics that show a clear relation to system performance are mainly the number of search terms and the number of hit conditons. As to the number of search terms it seems reasonable

that it should be positively related to output, number of relevant references and user satisfaction. The probability of retrieving a reference, relevant or irrelevant, is of course greater if the number of search terms is increased. This should hold true at least at a general level. However, if the search logic is narrow the effect of the number of search terms will be diminished. On the other hand the number of search terms ought to be negatively related to precision. That is, the probability of retrieving irrelevant references increases when the number of search terms increases.

The number of hit conditions seems to have a similar relationship to system performance and user satisfaction. However, it seems obvious that the number of hit conditons should be neither too low nor too high in order to get a good performance.

Another aspect of profiling is the amount of activity that is expended on making profiles. The only measure available here is the mean number of profile versions. The mean number of profile versions for users with low, medium and high satisfaction are respectively 2.5, 2.5 and 2.9. Thus an increase in the number of porfile versions seems to be positively related to user satisfaction.

Yet another aspect of profiling is whether users themselves have initiated profile changes by making contact with the IDC. Among users with low satisfaction 33 percent had initated such profile changes, 37 percent among those with medium satisfaction and 39 percent among those with high satisfaction had done so.

Generally, profiling activity seems to improve system performance. The efforts that could be expended on profiling were however limited during the test period.

8.5 User characteristics

Some aspects of how user characteristics affect the IR-process have already been mentioned. User knowledge about system functions was

shown to be positively related to system performance. User knowledge about system functions was assumed to be of importance for query formulation and profiling.

The user's subject area should be decisively related to the amount of information covered by the SSCI. We have found that the number of journals covered by the SSCI in user subject areas affects performance: The more journals covered in the subject area the greater the possibility of retrieving a relevant reference. Here we describe the relationship between subject area and satisfaction (Table 8.5).

Table 8.5 Usefulness of the SSCI in relation to subject area and number of journals covered by the SSCI. Figures indicate percentage of users with low, medium and high satisfaction.

	Usefulness of the SSCI			Number of jour-
Subject area:	Low	Medium	High	nals in SSCI
Business Administr.	54	31	15 (59)	50
Political Economy	67	27	16 (15)	78
Psychology	36	34	30 (58)	145
Education	21	23	56 (43)	84
Sociology	35	30	35 (46)	89
				446

From Table 8.5 we can conclude that the behavioural sciences have a notably greater satisfaction with the SSCI than economics and business administration. The SSCI-coverage in term of journals is also better for the behavioural sciences. An interesting result is that education has a relatively high degree of satisfaction although the SSCI-coverage is moderate. Perhaps the high degree of satisfaction among users in education is due to the fact that the documentalists at the IDC via the data-base ERIC are more accustomed in serving these users than others, as well as the fact that user in education themselves have a greater familiarity with computer-based IR-services. Educational research is furthermore a rather wide subject area utilizing scientific knowledge from various disciplines.

8.6 Inter-relations between some criteria of the IR-process

Up to now, we have studied the relationship between the different criteria of the IR-process and user satisfaction. The criteria used indicated various aspects of a specific phase that could be assumed to be of importance for the end result of the IR-process.

One important question that remains to be answered is in what way do different phases interact with each other during the IR-process. It would for example be of interest to know to what extent user knowledge of system functions affects the profiling activity in terms of profile/query correspondence, and the relationship between profile/query correspondence and user satisfaction. Since the IR-process has a logical time order between the different phases we can theoretically propose a causal model that describes the inter-relations between different phases in the following way:

Introduction $\Longrightarrow$ Query formulation $\Longrightarrow$ Profiling $\Longrightarrow$ System performance

This highly simplified causal model implies that: the introduction phase affects the query formulation phase, query formulation will have an effect on profiling and profiling will have an effect on system performance. This model is of course not empirically testable in this general form. In order to test a model that describes the inter-relations between phases of the IR-process we have to have more explicit definitions of the criteria involved, and connect the criteria into testable hypotheses.

To begin with we have to identify the main components of the process. The IR-process can be divided into two main parts. The introduction, query formulation and profiling phases have as their ultimate goal an adequate representation of the user's special interest area in the form of search profiles. This part involves the participation of two main components of an IR-system: users and documentalists. The end product of user/documentalist co-operation is search profiles.

The second part consists of the mechanized retrieval of references from the data-base. The main components involved here are search profiles and the data-base, the interaction of which leads to system performance. The inter-relation between these two main parts can be illustrated in the following way:

Starting with this preliminary sketch of the inter-relation between components of the IR-process we can define the relations at an empirical level using the variables or criteria we have studied in the previous chapters.

The first opportunity for interaction between users and documentalists occurred at the working seminars. The aim of these seminars was twofold. Firstly, users were taught the principles of computerized IR-service in order to give them some degree of interaction competence. Secondly, working seminars gave users an opportunity to discuss the first profile version with documentalists and to modify the profiles if necessary. If these two aims were achieved the working seminars were assumed to have a positive effect on profile/query correspondence.

Another factor that is of importance for profile/query correspondence is the amount of activity documentalists expended on profiling. A rough measure of documentalist activity is the number of profile versions. However, the number of porfile versions is also affected by initiatives from the users. These user initiatives are of two kinds, those made during the working seminars and those made after.

Looking at this first main part for the IR-process the main question to be answered is: What is the relative importance of user interaction competence and profiling activity for profile/query correspondence? In order to estimate the relative importance of these variables we

used a multivariate method developed by Coleman (1964). This method makes it possible to estimate the relationship between two variables while controlling for the influence of other variables.

As a measure of users' interaction competence we used the following item from the survey:

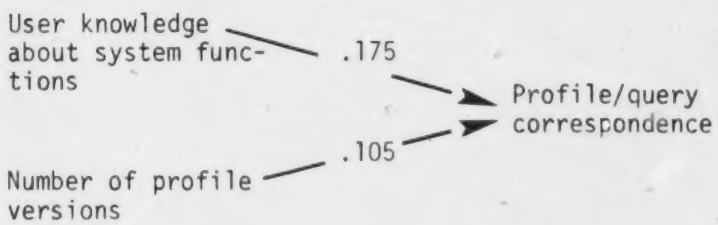
> "Do you feel that you have enough information about the SDI-service to be able to understand its function?"

Profile-query correspondence was indirectly measured by the following item:1)

> "Do you consider that the documentalists at the IDC received enough information about your special interest area to be able to make an adequate representation of your search profile?"

Profiling activity is measured by the number of profile versions. Fig. 8.1 shows that users' knowledge about system functions has a greater relative importance for profile/query correspondence than the number of profile versions.

Fig. 8.1 The relative effects of user knowledge about IR-system functions and number of profile versions towards profile/query correspondence.



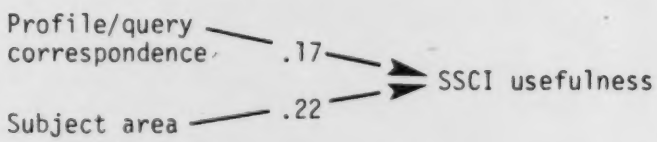
In the second part of the IR-process the interaction between search profile and data-base determines system performance. In previous

1) This item primarily measures whether or not the information delivered to the IDC was enough to permit a good profile/query correspondence. However, the answer to this question implies the user's satisfaction with the correspondence between his query and profile.

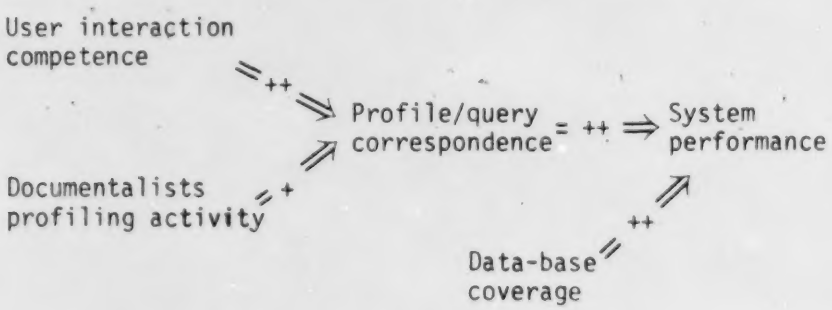
sections we have studied different characteristics of search profiles in relation to system performance. In this analysis we used the same item for profile/query correspondence as above in order to get an overall measure of profile quality. One of the aspects of the data-base that is crucial for system performance was found to be the users' subject area in relation to data-base coverage. Since subject area is closely correlated to coverage we used subject area to indicate the journal coverage in SSCI. [1] Finally, system performance is measured by users' overall estimation of SSCI usefulness.

In order to estimate the relative importance of profile/query correspondence and subject area for system performance we used the same multivariate method. Fig. 8.2 shows that both variables have an effect on system performance, but subject area has a somewhat greater relative importance (that is, the effect of being a behavioural scientist).

Fig. 8.2 The relative effects of subject area and profile/query correspondence on SSCI usefulness.



The models presented in Fig. 8.1 and 8.2 may be brought together to illustrate the inter-relationships between different criteria of the IR-process. Although not all empirical indicators used fully cover the theoretical variables we can propose the following model:



1) i.e. by using the dichotomy behavioural scientists - not behavioural scientists.

This model implies that user interaction competence has a somewhat stronger effect on profile/query correspondence than profiling activity. Data-base coverage and profile/query correspondence have an equal effect on system performance.

If this preliminary model holds true, it allows us to make assumptions regarding the effects of various practical actions for improving the service. Profile/query correspondence should be improved if users interaction competence is improved. This seems to be of greater importance for profile/query correspondence than the raising of more resources for profiling activity. This does not imply that resources for profiling are unnecessary, it merely indicates that user interaction competence is a more strategic factor, of greater relative importance. System performance cannot merely be improved by improving profile/query correspondence. Improvement of data-base coverage is of at least equal importance.

The models outlined in this section are however preliminary due to the fact that the quality of data used is limited. With a complete data collection for all phases of the IR-process, it would also have been possible to estimate the relative importance of each specific phase for performance. This is not possible here since query and profiling data particularly are incomplete.

IX SUMMARY AND CONCLUSIONS

The evaluation study described in the previous chapters contains a great number of results which indicate courses of practical action that would improve the service. These are discussed in this chapter and the practical implications are related to the different phases of the IR-process.

9.1 Improving the different phases of the IR-process

The introduction phase has a decisive influence on the phases that follow in the IR-process. The introduction phase involves the selection of the user group, which means that the limits for correspondence between data-base and user information needs are determined in this phase. The selection of a user group is a two-step process. Firstly, the IDC defines a potential user group, i.e. those persons or organizations that are to be offered the service. Secondly, the ultimate decision to accept the offer is left to the potential users. These decisions, which determine the final user group are of great importance for system performance. The rational basis for such decisions will therefore be crucial.

For rational decision-making, the IDC needs knowledge about general patterns of information needs and uses among potential user groups, just as the potential users who are offered the service need knowledge of its potentialities and limitations. In this context the situation was that very limited knowledge was available regarding information needs and uses in the social sciences. Nevertheless, the definition of a potential user group was not too difficult. However, the lack of earlier experience of social science information services gave rise to an underestimation of the total information need, which resulted in a stress situation, described in Chapter I.

The established user group is evenly distributed among the main disciplines. The only case of marked under-representation when the user group is compared to potential users is political economy.

The decision to participate should ideally be based on an adequate knowledge of the potentialities and limitations of the system. When this is not the case users might be misled into expectations of the system that cannot be met. As was noted in Chapter III many users seem to lack basic knowledge about computerized IR-processes. Such a knowledge is not only a necessary condition when users decide to use the IR-system or not, but also a strategic factor in the whole IR-process Related to this are our results from a study of technologists; when asked about their interest in a computerized information service a remarkably large number of respondents were undecided, thereby indicating their unsatisfactory decision situation.

Conclusion: *The introduction phase is crucial for system performance. Here, the user group is determined, the first contact is established, the limits for correspondence between data-base and user information needs are set and a base for the following phases is made.*
The evaluation study showed the vital role user knowledge about IR-processes has on system performance, but the importance of a thorough knowledge about user's information needs and uses has also been underlined.
Improvement of the introduction phase should consequently be based on increasing user knowledge about IR-system functions as well as augmenting the knowledge among the information service providers of the information needs and uses of the users.

Query formulation phase

The principal demand in the formulation of queries is that the users define their information needs so that correct interpretations of these are possible and easy to translate into profiles. This implies that users should employ the commonly used terminology of their interest area as it is found in the type of literature covered by the data-base. It also implies that some knowledge of profiling and the search processes used in the IR-system is beneficial. The study of the queries submitted

showed that query clarity and completeness were extremely unsatisfactory. The necessity for query clarity is even more pronounced when resources for profiling and user contact are limited. Since the SSCI only allows title search it is also important to have queries with a large number of search terms as a basis for profile construction.

Conclusion: *The SSCI-service could be improved by a combination of detailed requirements in the query forms, broader social science experience of documentalists and means of channels for closer contact between users and the IDC.*

Profiling phase

Looking at the profiling phase we found that the amount of profile activity in terms of profile versions constructed had a positive effect on system performance. This indicates that resources for profiling at the IDC should have a positive effect on system performance. One aspect of the profiling activity that was not measured in the evaluation study was the impact of differences between documentalists. However, the importance of special competence in the subject area is well known and could also be observed during our particpation at the IDC.

As to the profile characteristics studied, the number of search terms was found to have the most positive effect on performance in this phase. The importance of efforts from users and documentalists in order to find alternative representations of query meaning is thus corraborated.

Conclusions: *Increased resources for profiling and profile modification should improve system performance. The profiles in this title search system should be constructed with a greater number of search terms.*

Search phase

In the search phase, the coverage of the data-base for different subject areas had a clear impact on performance. The difference in performance

between disciplines was evident. While the result in the behavioural sciences was generally satisfactory, other disciplines such as economics had a rather low performance level. Although this result is largely consistent with differences in the coverage of the data-base, the number of journals should be enough to also satisfy many of the now dissatisfied customers, provided that the recommendations made above are implemented.

Conclusion: *The coverage of the data-base had a clear impact on system performance, but this impact is difficult to study in isolation because of the interaction with many other variables that are associated with discipline matters. The coverage of the data-base should allow a substantial increase in the number of satisfied users in many areas of study in the social sciences (The journal coverage is given in Appendix VI).*

Output

During the first test-period the performance rate (according to traditional measures of performance) has been slightly under the performance level reached in other and more established data-bases at the IDC. We found that there was a general tendency for low output, which in combination with a high frequency in the delivery of reference lists and the limited current awareness among users were possible grounds for dissatisfaction.

Conclusion: *Each reference list received should ideally contain a minimum number of relevant references. In order to make this more probable an increase in output is desirable at equal or better precision. This could be achieved by more infrequent mailing since a need for speed is not predominant in the social sciences.*

System performance

While output was somewhat low, the mean precision rate however was

quite satisfactory. Users' overall evaluation of SSCI:s usefulness showed that about 65 per cent of the users considered the SSCI to be useful. Only 35 per cent felt that the SSCI had been of little or no value during the test period.

Conclusion: *Generally, the performance of the SSCI is somewhat lower than the performance of other data-bases run at the IDC. However, taking the improvements suggested earlier into consideration, the performance of the SSCI-service could obviously be upgraded.*

9.2 Further marketing of the SSCI

In this report we have evaluated different phases of the IR-process during a test period with SSCI. A number of improvements have been proposed in order to remedy some of the initial problems encountered. One question that has not yet been dealt with is however whether the system should be available in the future or not. In order to deal with this question a number of other questions must be discussed.

Is there a need for computer-based information in the social sciences? In those studies quoted in Chapter I it was obvious that social scientists in general lack the means for retrieving references. Thus, it is obvious that a need for secondary sources does exist. Logically, the need for computerized IR-service is also probable. We cannot definitely say that in testing the SSCI among social scientists in Sweden, we have tested the general value of computerized IR-services. However, the users seems to appreciate the availability of this kind of information service. In the main survey the users were asked to state the desirability of future access to SSCI (Table 9.1).

Table 9.1 Percentage of users in discipline who considered future access to SSCI desirable.

	Percentage	n
Political Economy	100	9
Economic History	80	5
Education	71	42
Business Administration	70	44
Sociology	70	43
Statistics	70	10
Psychology	64	55
Cultural Geography	62	26
Political Science	59	22
Total number	175	269

About 65 per cent of the users thought that the SSCI would be desirable for them in the future. These users were rather evenly distributed according to subject area/discipline. At the end of the test-period in January 1974, 82 users (18% of the test group) wanted to continue their subscriptions with a half-yearly fee of SKr 100. Since 65 per cent of the users wanted to have access to the SSCI in the future, the 82 subscriptions represent a minority of possible SSCI users.

As to further marketing of the SSCI it should be obvious that the IDC has a future market in about 50 per cent of the users who participated. This gives us a rough number of 250 out of 500 who are interested in having future access to the SSCI. If we assume that in general about 65 per cent of Swedish social scientists are eager to have access to the SSCI, this would imply that about 750 of the potential users defined initially (1159) are possible future users of the SSCI.

Since the potential users defined initially are limited to social scientists at universities, a number of other potential user group can be found in special research insitutes, government agencies and industry.

9.3 The general value of SSCI for social science research and practice

As was noted in Chapter 1 there is a general lack of an effective

means for searching the literature in the social sciences; this was also indicated by the large number of initial requests for this service. It is also obvious that a small country like Sweden, where only about one per cent of the world's total research activities are performed, must be highly dependent on international research literature for keeping up with scientific developments and new techniques. In this context the effects of introducing a new current awareness service cannot be evaluated only from a consideration of the delivery of references to social science researchers. The value of scientific information is determined in research practice and ideally becomes manifest in raised cumulativity and better research results. Other possible effects of introducing a new literature service are that it might change the information seeking patterns and the level of current awareness among the users. Although effects of this kind are difficult to detect during a short period of time we have tried to obtain some indications of such effects by posing retrospective questions immediately after the 13-week test period.

SSCI was designed as a current awareness service. Then, if SSCI is of general value for social science the users knowledge of current research should improve. As Table 9.2 indicates a majority of the users felt that they had experienced at least some improvement of their knowledge about current research in their subject field.

Table 9.2 Question: »Do you think you have gained a better knowledge about the current research within your subject field through the SDI-service?»

	Per cent
Much better	2.7
Better	13.0
Slightly better	44.0
Unchanged	40.3
	(330)

The service based on SSCI obviously had some positive effects on the current awareness among users during the relatively short test period.

Many users, however, did not feel that way. The answers given in Table 9.2 do not only reflect the effects of the service based on SSCI, but also those effects due to the availability of original documents and library resources as well as the demand for current awareness among the users. A question of this kind thus partly "evaluates" not only the SSCI-service, but also library services and the users themselves in unknown proportions. Looking back at our earlier results we know from Chapter I that the demand for current awareness is limited among social scientists compared to other scientists. We have also shown that some users did not retrieve any relevant references and that those who did have not retrieved more than a third of the relevant references in original form. It was also found that about 50 per cent of the users had some difficulties in retrieving original documents. This indicates that the service ability of Swedish libraries has to be improved in order to be able to satisfy the increased demand for social science journals resulting from a computer-based literature service of this kind. The difficulties in retrieving original documents obviously are not solely dependent on the physical availability of journals, but also on the library skill among users and the interaction between users and librarians as to the procedures used for requests etc.

In measuring the general effects of SSCI usage we also asked the users whether they had changed their reading habits as a result of SSCI-usage (Table 9.3).

Table 9.3 Changes in reading journals. Per cent.

Increased reading	15.7
Unchanged reading	83.7
Decreased	0.6
	(327)

SSCI-usage thus seems to have stimulated the reading of social science journals among about 15 per cent of the users. We could not find any

corresponding decrease in the utilization of other information sources such as books.

Another effect of the use of SSCI might be that users are stimulated to read articles from journals not previously used (Table 9.4 shows that almost every user became acquainted with new journals in his area of interest.

Table 9.4 Question: »Have you become aquainted with journals of interest to you that you did not know of previosly?»

Number of journals	Per cent
None	0.6
1-2	41.4
3-4	27.7
5-10	21.7
11-	1.9
	(314)

Summing up this evaluation study we can conclude that the potential value of a computerized IR-service based on SSCI has been demonstrated in practice in spite of the problems encountered during a short introductory test period.

REFERENCES

Allen T J, Managing the Flow of Scientific and Technical Information, Dissertation, Massachusetts Inst of Techn, 1966

American Psychological Association, Reports of Project on Scientific Information Exchange in Psychology, Vol 1-3 (1963,-65,-69)

Coleman J S, Introduction to Mathematical Sociology, The Free Press of Glencoe, 1964

Garvey W D, Lin N, Nelson C E, A Comparison of Scientific Communication Behaviour of Social and Physical Scientists, Int Soc Sc J, Vol XXIII (1971) 2, pp 256-272

Goldberg A, Information Needs of Social Scientists and Ways of Meeting Them, Int Soc Sc J, XXIII (1971) 2, pp 236-243

Goodman L A, Kruskal W H , Measures of Association for Cross Classifications, J of the American Statistical Association, 49 (1954) dec, pp 732-764

Hjerppe R, Serviceability - A Meta-Model for Evaluating IR Serviceability, in Samuelsson K et al, Evaluation of Information Retrieval Systems and Functions, NORDFORKS - The Scandinavian Council for Applied Research, Nov 1974

Lancaster F W, Information Retrieval Systems, John Wiley & Sons Inc, New York 1968

Line M B, Information Requirements in the Social Sciences: Some Preliminary Considerations, J of Librarianship, 1 (1969), pp 1-19

Menzel A, Formal and Informal Satisfaction of the Information Requirements of Chemists, Columbia University, Bureau of Applied Social Research, June 1970

Rosenblom R S, Wolek W, Technology and Information Transfer, Division of Research Graduate School of Business Administration, Havard University, Boston 1970

Scheffler HH, et al, The Significance of Titles, Abstracts and Other Portions of Technical Documents for Information Retrieval, IEEE Transactions of Professional Communication, PC-17 (1974)

UKCIS, Research Report No. 2, Retrieval Experiments Based on Chemical Abstracts Condensates, Nottingham, 1974

Vickery B C, Information Systems, Butterworths, London, 1973

Zaltman G, Scientific, Recognition and Communication Behaviour in High Energy Physics, Unpublished Ph.D. Thesis, John Hopkins University, Baltimore, Maryland, June 1968 (P179890).

APPENDIX I

SDI-INFORMATION 1053

1:e bibliotekarie
Zofia Gluchowicz, 1125, EH

1973-03-26

Kostnadsfri litteraturbevakning i 13 veckor.

Vill Ni deltaga i ett utvärderingsexperiment som ger Er möjlighet att kostnadsfritt under 13 veckor följa upp litteraturen inom samhällsvetenskaperna?

Ni får kostnadsfritt varje vecka en lista över nypublicerade tidskriftsartiklar och rapporter och annan litteratur. Varje deltagare erhåller en lista som är individuell och anpassad exakt efter vars och ens intresseområde. Urvalet sker ur de 1000 mest citerade tidskrifterna inom samhällsvetenskaperna. Det enda Ni behöver göra är att beskriva Ert intresseområde för en dokumentalist vid KTH:s bibliotek, göra en enkel värdering av varje litteraturlista och slutligen besvara ett frågeformulär vid utvärderingstidens slut.

KTH:s bibliotek har idag 5 års erfarenhet av datorbaserad dokumentation. Hittills har denna verksamhet bedrivits inom den naturvetenskapliga sektorn med mycket gott resultat. KTH:s SDI service, som den kallas, söker idag igenom ca 1,3 miljoner referenser per år. Det motsvarar ca 10 000 olika tidskrifter. I dessa tidskrifter finns en hel del information som är av intresse för samhällsvetare, men hittills har det inte funnits någon databas speciellt gjord för samhällsvetenskaperna.

KTH:s bibliotek har nu fått ett erbjudande från Institute for Scientific Information i USA att som prov och för utvärdering erhålla deras databas Social Sciences Citation Index - SSCI - under 13 veckor.

KTH:s bibliotek är alltså intresserat av att undersöka om det i Sverige finns ett behov av en datorbaserad litteraturbevakning bland samhällsvetarna och om detta behov kan täckas av den erbjudna databasen i kombination med de bandsystem som redan finns och som behandlas på KTH:s bibliotek.

Det kan nämnas att KTH:s bibliotek sedan 1971 i samarbete med statens psykologisk-pedagogiska bibliotek behandlar en databas som kallas ERIC, Educational Resources Information Center, som innehåller material för pedagoger.

Som projektledare står 1:e bibliotekarie Zofia Gluchowicz till förfogande för närmare upplysningar. Tel: 08/23 65 20 ankn 1125

Om Ni är intresserad av att deltaga så var vänlig och fyll i bifogade formulär och sänd det till: KTH Biblioteket, Dokumentationstjänst, 100 44 STOCKHOLM.

SDI
Selektiv Delgivning av Information
Kungl Tekniska Högskolans Bibliotek
Dokumentationstjänst

APPENDIX II

NAMN.....

INSTITUTION/ORGANISATION.....

ADRESS.....

.....TEL.....

Beskrivning av intresseområdet i klartext i några meningar:

1) Estimation of interdependent systems with auto-correlated residuals. Interdependent systems non-linear in parameters and/or variables. Non-linear maximum-likelihood estimation. Generalized least squares.

2) Econometric macro-models. Interdependent systems applied on national basis, Transition from economic theories and theory-systems to econometric models. Econometric models formulated from marxian theories.

APPENDIX III:1

```
  NAME      : SDIX072                                        PAGE 01
  DATA BASE : SSCI                  VERSION : 001

* RANK      : 90                    MODIFICATION DATE : 73-05-30
* MAX. REFS : 0100                  CREATION DATE     : 73-05-30

* COMMENTS  : ESTIMATION OF INTERDEPENDENT SYSTEMS

* LOGIC     : A*B*((C+E)*D +C*E +H) +F*G

  GROUP NO TYPE WEIGHT CUM   TERM

*   A   01 TITL   +02       * INTERDEPENDEN*

*   B   01 TITL   +02       * SYSTEM*

*   C   01 TITL   +02       * ESTIMAT*

*   D   01 TITL   +02       * AUTOCORRELAT*
*   D   02 TITL   +02       * AUTO CORRELAT*
*   D   03 TITL   +02       * RESIDU*
*   D   04 TITL   +02       * MAXIMUM LIKELIHOOD*
*   D   05 TITL   +02       * LEAST SQUARE*
*   D   06 TITL   +02       * PARAMETER*
*   D   07 TITL   +02       * VARIABLE*

*   E   01 TITL   +02       * NON LINEAR*

*   F   01 TITL   +02       * ECONOMETRIC*

*   G   01 TITL   +02       * MACRO*
*   G   02 TITL   +02       * MARX*
*   G   03 TITL   +02       * MODEL*
*   G   04 TITL   +02       * THEOR*
*   G   05 TITL   +02       * TRANSIT*

*   H   01 TITL   +02       * NATIONAL*

  TOTAL NO. OF TERMS : 0018
```

APPENDIX III:2

NAME : SDIX072 PAGE 01
DATA BASE : SSCI VERSION : 002

RANK : 90 MODIFICATION DATE : 73-06-28
MAX. REFS : 0100 CREATION DATE : 73-05-30

COMMENTS : ESTIMATION OF INTERDEPENDENT SYSTEMS

LOGIC : A*B*((C+E)*D +C*E +H) +F*G

	GROUP	NO	TYPE	WEIGHT	CUM	TERM
	A	01	TITL	+02		* INTERDEPENDEN*
	B	01	TITL	+02		* SYSTEM*
	C	01	TITL	+02		* ESTIMAT*
	D	01	TITL	+02		* AUTOCORRELAT*
	D	02	TITL	+02		* AUTO CORRELAT*
	D	03	TITL	+02		* RESIDU*
	D	04	TITL	+02		* MAXIMUM LIKELIHOOD*
	D	05	TITL	+02		* LEAST SQUARE*
	D	06	TITL	+02		* PARAMETER*
	D	07	TITL	+02		* VARIABLE*
*	D	08	TITL	+02		* DISTRIBUTED LAG*
*	D	09	TITL	+02		* AUTOREGRESS*
*	D	10	TITL	+02		* GLS *
*	D	11	TITL	+02		* TWO STEP*
	E	01	TITL	+02		* NON LINEAR*
	F	01	TITL	+02		* ECONOMETRIC*
	G	01	TITL	+02		* MACRO*
	G	02	TITL	+02		* MARX*
	G	03	TITL	+02		* MODEL*
	G	04	TITL	+02		* THEOR*
	G	05	TITL	+02		* TRANSIT*
	H	01	TITL	+02		* NATIONAL*

TOTAL NO. OF TERMS : 0022

APPENDIX III:3

```
   NAME       : SDIX072                                        PAGE 01
   DATA BASE  : SSCI                    VERSION : 003

   RANK       : 90                      MODIFICATION DATE : 73-09-05
   MAX. REFS  : 0100                    CREATION DATE     : 73-05-30

   COMMENTS   : ESTIMATION OF INTERDEPENDENT SYSTEMS

*  LOGIC      : A*B*(D+H)+C*E+F*G

   GROUP NO TYPE WEIGHT CUM   TERM

     A   01 TITL   +02       * INTERDEPENDEN*
*    A   02 TITL   +02       * SIMULTANEOUS*
*    A   03 TITL   +02       * MULTIRELATIONAL*

     B   01 TITL   +02       * SYSTEM*
*    B   02 TITL   +02       * EQUATION*

     C   01 TITL   +02       * ESTIMAT*
*    C   02 TITL   +02       * GLS *
*    C   03 TITL   +02       * TWO STEP*
*    C   04 TITL   +02       * MAXIMUM LIKELIHOOD*
*    C   05 TITL   +02       * LEAST SQUARE*

     D   01 TITL   +02       * AUTOCORRELAT*
     D   02 TITL   +02       * AUTO CORRELAT*
     D   03 TITL   +02       * RESIDU*
     D   04 TITL   +02       * PARAMETER*
     D   05 TITL   +02       * VARIABLE*
     D   06 TITL   +02       * DISTRIBUTED LAG*
     D   07 TITL   +02       * AUTOREGRESS*
*    D   08 TITL   +02       * SPECIFICATION*
*    D   09 TITL   +02       * DISTURBANCES*
*    D   10 TITL   +02       * MOVING AVERAGE*

     E   01 TITL   +02       * NON LINEAR*

     F   01 TITL   +02       * ECONOMETRIC*

     G   01 TITL   +02       * MACRO*
     G   02 TITL   +02       * MARX*
     G   03 TITL   +02       * MODEL*
     G   04 TITL   +02       * THEOR*
     G   05 TITL   +02       * TRANSIT*

     H   01 TITL   +02       * NATIONAL*

   TOTAL NO. OF TERMS : 0028
```

APPENDIX III:4

NAME : SDIX072 PAGE 01
DATA BASE : SSCI VERSION : 004

RANK : 90 MODIFICATION DATE : 73-11-06
MAX. REFS : 0100 CREATION DATE : 73-05-30

COMMENTS : ESTIMATION OF INTERDEPENDENT SYSTEMS

LOGIC : A*B*(D+H)+C*E+F*G

	GROUP	NO	TYPE	WEIGHT	CUM	TERM
	A	01	TITL	+02		* INTERDEPENDEN*
	A	02	TITL	+02		* SIMULTANEOUS*
	A	03	TITL	+02		* MULTIRELATIONAL*
	B	01	TITL	+02		* SYSTEM*
	B	02	TITL	+02		* EQUATION*
	C	01	TITL	+02		* ESTIMAT*
	C	02	TITL	+02		* GLS *
	C	03	TITL	+02		* TWO STEP*
	C	04	TITL	+02		* MAXIMUM LIKELIHOOD*
	C	05	TITL	+02		* LEAST SQUARE*
	D	01	TITL	+02		* AUTOCORRELAT*
	D	02	TITL	+02		* AUTO CORRELAT*
	D	03	TITL	+02		* RESIDU*
	D	04	TITL	+02		* PARAMETER*
	D	05	TITL	+02		* VARIABLE*
	D	06	TITL	+02		* DISTRIBUTED LAG*
	D	07	TITL	+02		* AUTOREGRESS*
	D	08	TITL	+02		* SPECIFICATION*
	D	09	TITL	+02		* DISTURBANCES*
	D	10	TITL	+02		* MOVING AVERAGE*
	E	01	TITL	+02		* NON LINEAR*
*	F	01	TITL	+02		* ECONOMETRIC *
	G	01	TITL	+02		* MACRO*
	G	02	TITL	+02		* MARX*
	G	03	TITL	+02		* MODEL*
	G	04	TITL	+02		* THEOR*
	G	05	TITL	+02		* TRANSIT*
	H	01	TITL	+02		* NATIONAL*

TOTAL NO. OF TERMS : 0028

APPENDIX IV

SDI
Selektiv Delgivning av Information
Kungl Tekniska Högskolans Bibliotek
Dokumentationstjänst

LITTERATURLISTA

Datum 73-12-06
Databas SSCI
Antal ref 2

Sökprofil SDIX072

SURVEY OF RECENT CANADIAN MACRO-ECONOMETRIC MODELS 1 S
TSURUMI H
CAN J ECON VOL 6 1973 N 3 P 409 R 047
WEIGHT=8.00 * MODEL* MACRO* ECONOMETRIC *

ESTIMATION AND STRUCTURE OF SECOND ECONOMETRIC 2 S
MODEL OF HUNGARY
HALABUK L
MATEKON VOL 9 1973 N 4 P 43 R 024
WEIGHT=4.00 * MODEL* ECONOMETRIC * ESTIMAT*

APPENDIX V

Selektiv Delgivning av Information

Kungl Tekniska Högskolans Bibliotek
Dokumentationstjänst

INTRESSEVÄRDERING Formulär

litteraturlistans datum 73-12-06

databas

sökprofil SDI X072

beställare kontaktperson

företag/institution

adress

postadress

Kryssmarkera för varje referens i tillämplig kolumn.

1 Av omedelbart intresse. (Drag en ring runt referensnumret, om kopia av artikeln önskas.)

2 Av intresse, men har redan en kopia. (Har läst artikeln.)

3 Av sekundärt intresse.

4 Kan ej avgöra om artikeln är av intresse, då citatet ej är nog detaljerat.

5 Är ej av intresse, då materialet ej överensstämmer med den beskrivning av mitt intresseområde som jag givit. Irrelevant.

6 Är ej av intresse, då mitt intresseområde har förändrats sedan jag beskrev det för Er.

Pris för kopior:
Kontoinnehavare 50 öre sidan.
Studerande och lärare vid KTH 50 öre sidan.
Övriga 75 öre sidan.

Övriga synpunkter

Referensnummer	1	2	3	4	5	6
01	X					
02	X					
03						
04						
05						
06						
07						
08						
09						
10						
11						
12						
13						
14						
15						
16						